40

ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XL NUMBER 1 MARCH 1994

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

The Alberta Journal of Educational Research

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
University of Alberta

George Buck
University of Alberta

James Chapman
Massey University

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Linda LaRocque
University of Alberta

Stewart J.H. McCann
University College of Cape Breton

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Sharon Robertson
University of Calgary

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Graham Fishburne
Robert H. Short
Fern Snart

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Publication Services
4-116 Education Centre North
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0390

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XL, Number 1, March 1994

Judy Cameron	**1**	Editorial *AJER*: Forty Years
George H. Buck	**3**	Herbert T. Coutts and the Origins, Early Development, and Possible Future Directions of the *Alberta Journal of Educational Research*
Catherine D. Krull	**7**	Level of Education, Sexual Promiscuity, and AIDS
Keith D. Walker *R.J. Carney*	**21**	Notions of "Ethical" Among Senior Educational Leaders
H.W. Hodysh	**35**	History of Education and the Rite of Passage to Teaching: The Alberta Experience 1893-1945
Eileen Wood *Teena Willoughby* *Violet Kaspar* *Tracey Idle*	**57**	Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations
Joseph Blase *Jo Roberts*	**67**	The Micropolitics of Teacher Work Involvement: Effective Principals' Impacts on Teachers
George Buck *Birendra K. Sinha* *Terry W. Belke*	**95**	Book Reviews

ISSN 0002-4805

Acknowledgment

The following people have helped to maintain the quality of AJER by reviewing manuscripts during the past year. Their contributions are greatly appreciated.

Armstrong, K.
Belke, T.
Buck, G.
Burnaby, B.
Cameron, D.J.
Clandinin, J.
Coleman, P.
Connelly, M.
Derwing, T.
Ellis, J.
Evans, P.
Everett-Turner, L.
Fris, J.
Fullan, M.
Gaskell, P.J.
Geddis, A.
Gillette, M.
Gustafson, B.
Harley, B.
Holdaway, E.A.
Hunka, S.
Hutchinson, N.
Huvasberger, M.
Kelly, U.
Levin, B.
Marshall, D.
Massing, D.
McIntosh, G.
Miklos, E.
Musella, D.
O'Shea, T.J.
Oberg, A.
Patterson, R.S.
Peters, F.
Reynolds, C.
Rees, R.
Richards, D.
Ross-Epp, J.
Ryan, H.
Samiroden, W.
Sherrill, J.M.
Sigurdson, S.E.
Stewart, D.
Sullivan, K.
Svenson, L.
Symons,S.
Thiessen, E.
Titley, B.
Townsend, R.
Vallance, J.
Vaselenak, L.
Vibert, A.
Walker, L.
Williams, T.
Wilson, D.N.
Wilson, L. J.
Wright, P.

Editorial

I would like to take this opportunity to introduce myself as the new editor of *AJER*. I am an assistant professor in the Department of Adult, Career and Technology Education at the University of Alberta. I completed my doctoral studies in educational psychology at the University of Alberta in 1992. My interests include learning and motivation, behavior analysis, second language teaching and learning, and research methodology.

A special thanks to Dr. Rob Short, my predecessor, for his guidance and help in my new role as editor. Rob served as the editor for the past two years. During his term three special issues of *AJER* appeared. In December 1991, *AJER* devoted an issue to the 50th anniversary of the Faculty of Education at the University of Alberta; in March 1993, under guest editors Jacques Désautels and Colette Dufresne-Tassé, francophone educational research in Quebec was featured; and in June 1993 Nelly McEwen edited a volume that included papers presented at the Charlottetown symposium on the Educational Quality Indicators (EQI) initiative. In addition, under Dr. Short's guidance, *AJER* began to publish French abstracts of each article. *AJER* continues to be a scholarly educational publication that addresses many interests and deals with a wide range of educational issues. The time, energy, and initiative Rob took as editor is greatly appreciated.

In each March issue *AJER* recognizes the contributions of the past year's reviewers. Their critiques constitute an essential part of the publication process and I thank them for a job well done. Finally, the advisory committee and I would like to welcome 10 new consulting editors: Terry Belke, George Buck, John Connors, Samuel Deitz, Sharon Haggerty, Antoinette Oberg, Ruth Rees, Rob Short, Kelleen Toohey, and Robert Wilson. I look forward to working with all of you to produce a quality journal that continues to be an important publication in the field of education.

AJER: Forty Years

This issue of the *Alberta Journal of Educational Research* marks the beginning of its 40th year in publication. Volume 1, Number 1 was published in March 1955. We are proud to share the 40th anniversary of *AJER* with our readers and authors who have contributed their support to the mission of this journal. In this issue the article by George Buck is based on an interview with Herbert T. Coutts who was involved in establishing the journal. The article provides an overview of the origins of *AJER* and highlights some of the developments that have taken place during four decades of publication.

The *Alberta Journal of Educational Research* is the oldest educational journal in Canada. Over the years *AJER* has enabled educational scholars to communicate and interact with one another. In the early years, *AJER* provided a forum for academics, students, teachers, and administrators from Alberta. As time passed, contributors and subscribers came to reflect a broader audience. Today

authors and readers come from across Canada and from many other parts of the world. *AJER* has gained a reputation as a broad-based, high quality educational journal of interest to an international community. I thank the many scholars, researchers, and practitioners who have published their work in *AJER*. Past editors, editorial board members, and those who refereed the thousands of articles submitted to the journal deserve a special thank you. These individuals have rendered an important service to the field. I am proud to take over as editor on this 40th anniversary of *AJER*.

Like all editors, I have my own point of view about educational research methods, style, and content of manuscripts. For 40 years *AJER* has published a wide range of studies with varying methodologies in most (if not all) areas of education, and I intend to uphold that tradition and build on the journal's eclecticism as a strength. Today there are many specialized journals in education; such journals may be difficult to access when one is not familiar with the specialty area. In contrast, *AJER* provides an opportunity for authors in specialized areas to communicate to a diverse readership. This places demands on contributors; specific issues should be made relevant to the larger educational community. Terms and methods must be clearly and simply outlined in as jargon-free language as possible. Authors should explain what the generality and practical importance of the article is.

Currently many educational researchers appear to spend a good deal of time and energy squabbling about how to do research. Arguments concerning quantitative versus qualitative, statistical versus single-subject design, historical/philosophical analysis versus field observation or experiment, and so on, in my view, occupy a disproportionate amount of academic educators' time. From my perspective, these squabbles are unnecessary, because different positions, methods, or approaches are appropriate for different analyses or questions. *AJER* will accept submissions from any and all who choose a method, position, or analysis that is appropriate to the inquiry of the paper.

AJER welcomes research papers that use statistical, qualitative, or single-subject designs. Articles that evaluate educational questions within a broader historical perspective are also encouraged. Such a context allows us to see whether education as a practice is progressing or simply cycling. Philosophical analysis is equally important because it often drives educational research and practice. Especially welcome are reviews that integrate and clarify educational issues. What I hope will distinguish *AJER* from other journals is clear, jargon-free articles from diverse areas that point to educational implications for a broad readership.

Judy Cameron

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 3-6*

George H. Buck
University of Alberta

Herbert T. Coutts and the Origins, Early Development, and Possible Future Directions of the *Alberta Journal of Educational Research*

In celebrating the 40th anniversary of the *Alberta Journal of Educational Research,* it is appropriate to investigate the reasons of why and how it came to be, the factors contributing to its establishment, and why it evolved as it did. When researching the origins and development of a given phenomenon, one is usually able to locate primary source material that lists significant dates, names, and events. Rarely, however, is one fortunate enough to find contributors to that phenomenon who are able to provide further elucidation, such as why the phenomenon occurred at that time and why particular individuals were associated with it. We are most fortunate in that Dr. Herbert T. (Pete) Coutts, the third Dean of the Faculty of Education, University of Alberta, is available and has helped by providing information supplementary to the surviving accounts of the beginnings of the *Alberta Journal of Educational Research (AJER).* Through the combination of primary source material and Dr. Coutts' recollection we may gain a more complete perspective of the origins and development of the *AJER.*

The year 1955 was significant not only for the establishment of the *AJER,* but also for the appointment of Dr. Coutts as Dean. The year also marked 13 years from the establishment of the Faculty of Education, 10 years since the University of Alberta became the sole body responsible for teacher training in Alberta, and nine years since H.T. Coutts joined the Faculty. Dr. Coutts notes that the Faculty of Education was indeed a fledgling body of staff from disparate backgrounds, and that the Faculty was trying to gain credibility with the other faculties of the university. Coutts (1993) states, "Science people in particular felt that we had no research or interest in pursuing it. Although research was ongoing both in the schools and at the university, there were limited outlets for it, and none of these were centered at the university" (Personal interview, September 24). Coutts' statement parallels similar observations made by the first Dean of the Faculty of Education M.E. LaZerte (1951) and later observations by Dunlop (1953). Although publications such as the *ATA Magazine* and the *Alberta School Trustee* had for many years provided some local outlets for reporting educational research done in Alberta, neither publication had evolved into a scholarly journal, and neither was dedicated to reporting research primarily.

George Buck is a lecturer in the Department of Educational Psychology. He began his career as a teacher and includes school building design and its relationship to learning and instructional theory among his areas of interest.

Although there were many established scholarly educational journals in North America by this time, most were based in the United States, and even those based in Canada, according to Dunlop (1953), were not "devoted entirely to the reporting of the results of educational research" (p. 42). Coutts (1993) states, "much of the energy and direction for the creation of an educational research journal for Alberta came from [George Murray] 'Pat' Dunlop [Head of the Division of Educational Psychology at the University of Alberta, Edmonton Campus]" (Personal interview, September 24). From the standpoint of professors in the Faculty of Education, Coutts notes another factor that provided impetus for the establishment of a scholarly journal, "continuation and promotion at the university was in large part determined by the adage *publish or perish.*"

Dunlop (1954; 1955) notes that by the beginning of 1954 the need for a scholarly outlet for the dissemination of educational research findings in Alberta was recognized by three provincial bodies besides the Faculty of Education. These groups were the Alberta Department of Education (now designated as Alberta Education), the Alberta School Trustees' Association, the Alberta Teachers' Association, and the Alberta Federation of Home and School Associations. With such a wide range of support from provincial organizations it was believed that adequate funding could be secured for the publication of a journal to be known as the *Alberta Journal of Educational Research*. To this end two committees were established in February 1954 (Dunlop, 1954).

The first committee, based solely in the Faculty of Education, was called the Faculty of Education Research Committee. Its members comprised the Dean as Chairman, the Heads of the three divisions of the faculty at that time (Coutts was Head of the Division of Secondary Education), the Director of Summer Session, the Director and Assistant Director of Research, and the Editor of the proposed journal (Dunlop, 1954). The Research Committee was responsible for the publication of the journal, as well as for the supervision of sponsored research projects (Dunlop, 1955).

The second committee was the Alberta Advisory Committee on Educational Research, which oversaw the work of the first committee and which also provided financial assistance to selected research projects. This committee, which was also chaired by the Dean of Education, consisted of representatives from the other provincial bodies mentioned above, and the President of the University of Alberta, ex officio (Dunlop, 1955). The Alberta Advisory Committee on Educational Research was, therefore, the more influential of the two committees. Although Coutts was already a member of the Research Committee, he did not assume the Chair of the Advisory Committee until his installation as Dean, September 1, 1955, following the retirement of his predecessor, Dr. Herbert E. Smith.

Coutts (1993) recalls that Smith, "was truly a gentleman, with a knack of seeing the whole picture and getting you to see it too" (Personal interview, September 24). With this bearing, Smith helped focus the direction of the individuals on the Advisory Committee in spite of the varied interests that they represented. Under Smith's leadership and guidance, the first issue of the *Alberta Journal of Educational Research* was published in March 1955 (Dunlop, 1955). The first editor was Harold S. Baker, who had succeeded Coutts as Head

of the Division of Secondary Education (Coutts, 1979). The direction set by Smith was continued by Coutts when he became dean (Coutts, 1955).

It should not be taken for granted that the locus of the *AJER* was always considered to be the University of Alberta. Dunlop (1955) states, "The centering of the research program [and the *AJER*] in the University was justified by the concentration there of trained staff, graduate student research workers, offices, and the largest library on education in the province" (p. 21). After becoming dean and assuming the chair of the Advisory Committee, Coutts recalls that he soon discovered a situation that threatened to move the *AJER* away from the university, and in all likelihood alter the journal's direction from a purely scholarly base. Coutts (1993) states,

> I perceived that some members of the Advisory Committee, representing the desires of their organizations, felt that control of the Journal should be within their particular group, not the Faculty of Education. I was opposed to this, and I felt that a strong measure should be taken to prevent such a takeover from occurring. It is for this reason that I held the copyright for the Journal personally for a number of years. (Personal interview, September 24)

Moreover, Coutts also notes, "It was further thought by me, 'Pat' Dunlop and some others, that if we could keep the Journal based in the Faculty, then the quality of articles could be kept to a high level, and this would be seen favorably by granting agencies" (Personal interview, September 24). The strategy was sound, for Coutts and Black (1962) report that in 1956, "the Carnegie Corporation of New York announced the award of $50,000 to the University of Alberta as a stimulus to its [Faculty of Education] research program" (p. 6). A portion of this award went to the maintenance of the *AJER*.

Coutts and Black (1962) report that the Carnegie grant enabled the scope of the *AJER* to be expanded. The authors note that the Journal, "has maintained the high standard set for it and has become internationally known" (p. 6). The fact that a local research-based journal gained international repute in under 10 years of publication was surprising to some individuals. The authors note, "the acceptance of this publication, carrying as it does the news of research to a public heretofore not reached, has exceeded expectations" (p. 6). It was apparent that the *Alberta Journal of Educational Research*, had evolved into a scholarly journal with a wide readership.

Reflections on the Present and Future of the Journal

It has been 21 years since H.T. Coutts was Dean of Education and has had an active role in the Faculty of Education. Nevertheless, his interest in developments and possible future directions in the field of education remains keen. Dr. Coutts (1993) notes, "the continued existence of the *Alberta Journal of Educational Research* shows the rest of the university [of Alberta] as well as other institutions that research in education, especially that based in Canada, is very much alive and is being carried out at high standards" (Personal interview, September 24). When asked about contemporary criticism regarding educational research, namely, that it is mediocre and of little practical use, Coutts replied,

> Never before in this province and likely this country have teachers been so well prepared and never has there been so much interest in educational research. We

> and our graduates have nothing to be ashamed of. If anything, we have failed to convince individuals outside our profession that what we do is more than the basic job training as provided by the normal schools so many years ago. (Personal interview, September 24)

Although the *Alberta Journal of Educational Research* has survived and grown in the past 40 years, Coutts believes that its continued existence will depend on how much our profession will defend itself and its goals. Coutts (1993) states, "Bad press and divisiveness should be fought" (Personal interview, September 24). As Dean of Education, Coutts notes that he strove to achieve three goals that facilitated both the survival of education as a separate discipline and educational research as a viable enterprise,

> First, I built up contacts with other faculties, so that each understood what the other was doing and that we were all trying to achieve the same ends: the best possible education for our students. Second, I attempted to build up a staff where each member possessed the highest qualifications. Third, I emphasized the importance both of research and teaching, and that neither one was more important than the other. (Personal interview, September 24)

Part of the strategy to achieve these three aforementioned goals was the *Alberta Journal of Educational Research*. Coutts (Hodysh & McIntosh, 1982) states,

> The strongest supporters one can have are those who believe that what we are doing is important to them. When it comes to fighting for budgets and dollars that are also being fought for by others, it is helpful to have sympathetic and powerful allies. (p. 179)

Although Dr. Coutts hopes that the *Alberta Journal of Educational Research* continues to be published well into the future, he cautions that if educators lose their sense of professional pride, allow current programs to be decentralized to community colleges, and consider educational research to be appropriate only if it is congruent with current popular interests, then,

> we are likely to regress to a more primitive and inferior form of teacher education, similar to what existed in much of the continent [North America] before the Second World War. If this should happen, then much of what was fought for and accomplished by individuals such as LaZerte, Smith, Newland, and Dunlop and what has been accomplished through the publication of the *Alberta Journal of Educational Research* will be lost. We must maintain the wisdom, foresight and strength to resist such regressive steps. (Personal interview, September 24, 1993)

References

Coutts, H.T. (1955). Growing pains in education. *Alberta School Trustee, 25*(11), 10-12.

Coutts, H.T. (1979). *Yesterday: Tomorrow's dream and today's vision*. Edmonton, AB: Unpublished speech presented on the occasion of the dedication of the Herbert T. Coutts library.

Coutts, H.T., & Black, D.B. (1962). The Carnegie-University of Alberta joint project on educational research: A report after five years. *The Alberta School Trustee, 32*(6), 6-9.

Dunlop, G.M. (1953). Research in educational administration and supervision. *Canadian Education, 8*(4), 41-44.

Dunlop, G.M. (1954). Educational Research in Alberta. *Canadian Education, 9*(2), 18-25.

Dunlop, G.M. (1955). The Alberta advisory committee on educational research. *Alberta School Trustee, 25*(2), 21.

Hodysh, H.W., & McIntosh, R.G. (1982). Conversation with a dean: The life and times of H.T. Coutts. *Challenge in Educational Administration: The CSA Bulletin, 21*(4).

LaZerte, M.E. (1951). Presidential address. *Canadian Education, 7*(1), 20-29.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 7-20*

Catherine D. Krull
University of Alberta

Level of Education, Sexual Promiscuity, and AIDS

A disturbing development in the past decade has been the dramatic rise in the incidence of AIDS (acquired immunodeficiency syndrome). One popular view is that people with higher educational attainment are at lower risk for AIDS and other sexually transmitted diseases (STDs) than those with less schooling. The present study, using data from a national survey of the United States, reports the findings from two studies that indicate the popular view is incorrect. The results from Study 1 indicated that educational attainment had a statistically significant indirect effect on promiscuous sexual behavior through its impact on liberal sexual attitudes. Thus educated people have more liberal dispositions toward premarital sex than those with less schooling and this attitude promotes frequent sex with multiple partners. Study 2 focused on the implications of high-risk sexual practices for the incidence of AIDS in a population. It was found that the incidence of AIDS in a region is partly explained by the educational attainment, liberal sexual orientation, and the promiscuous sexual practices of the population. Overall, the results indicated that the level of education in a population indirectly increased the incidence of AIDS. Finally, the implications for educational policy are discussed.

On a témoigné depuis la dernière décenie, une croissance draconienne du SIDA (syndrome d'immunodéficience acquise). Il existe une opinion populaire qui cite que les gens ayant un niveau d'éducation plus élevé risquent moins de contracter le SIDA ainsi que d'autres maladies transmises sexuellement (STM) que ceux qui ont moins d'éducation. Cette étude utilise des données d'un questionnaire national provenant des Etats-Unis pour présenter les recherches de deux études qui indiquent que cette opinion populaire est incorrecte. Les résultats de la première étude indiquent que le niveau d'éducation a une signification statistique indirecte sur des comportements de promiscuité sexuelle par son influence sur les attitudes et les mœurs sexuelles libérales. En somme, les gens éduqués ont une attitude envers les relations prénuptiales beaucoup plus libérale que ceux qui possèdent une éducation moins formelle et que ces attitudes promouvoient plus de contacts sexuels avec plus de partenaires. La deuxième étude vise l'implication des comportements sexuels à risques élevés pour la présence du SIDA dans une population. On a prouvé que le taux de cas du SIDA dans une région est en partie expliqué par le niveau d'éducation de cette population, de son orientation sexuelle libérale et de sa pratique de promiscuité sexuelle. En somme, les résultats de ces études indiquent que le niveau de l'éducation d'une population était la cause indirecte du nombre plus élevé de cas du SIDA. On discute en dernier lieu, des implications de la politique de l'éducation.

A disturbing development in the past decade has been the dramatic rise in the incidence of AIDS (acquired immunodeficiency syndrome). One popular view is that individuals with higher educational attainment are at less risk for the disease. This view assumes that people with more schooling have formed values and attitudes that make them less likely to engage in behaviors that are associated with contracting the disease. The present study examines the im-

Catherine Krull is a doctoral student with the Department of Sociology. Her interests include social demography, education, and gender issues.

pact of educational attainment on the sexual attitudes and behaviors of unmarried adults in the United States. The second part of this study examines the effects of educational attainment on the incidence of AIDS in the population.

More than 234,000 Americans have contracted AIDS since 1981 and over 67% of them have since died (Centers for Disease Control, 1992, p. 5). The Centers for Disease Control report that the leading cause of death in the United States among men and women under the age of 45 years and among children between 1 and 5 years of age is the human immunodeficiency virus (HIV), which is believed to cause the acquired immunodeficiency syndrome (AIDS). Recent studies indicate that HIV rates among teenagers are increasing. "Three out of every 10,000 teenagers who applied for military service in the United States between 1985 and 1989 tested positive for infection with the human immunodeficiency virus (HIV)" (Burke, 1990, p. 2074). The sex ratio of reported AIDS cases among teenagers was close to unity (the rates were equally high for both males and females), but differed significantly by geographic region. This is in sharp contrast to the adult male-to-female ratio of reported AIDS cases, which was 8:1 in 1991 (Centers for Disease Control and Prevention, 1993, p. 11). However, male-to-female ratios are expected to decrease as HIV/AIDS continues to spread through heterosexual intercourse (United Nations, 1991).

It is well known that AIDS is preventable. Health education has been suggested as a means of changing attitudes and behaviors that are associated with the transmission of AIDS. The focus of health education and health policy has been to provide information on HIV/AIDS to targeted high risk groups. In this war of health education against AIDS, people with higher levels of formal schooling are seldom targeted as a group at risk for the disease. The belief is that schooling by itself provides the knowledge and motivation to avoid the risky behaviors associated with HIV transmission. For example, people with more schooling are expected to acquire attitudes and values that make them less likely to engage in sexually promiscuous behavior. Although this view is popular, Robertson (1989) argues that general education does not protect people from sexual promiscuity and AIDS.

> When the AIDS virus threatens lives, people look to education as the way to prevent the disease from spreading.... [However,] education alone has had little impact on people's unsafe sexual practices. The "let's-solve-it-through-the-schools" approach is based on the fallacy that students learn what they are taught, remember what they learn, and behave accordingly. The dismal fact of the matter is that ... even when students do learn and remember some of the values that the school teaches—such as those concerning sex ... —there is no guarantee that their behavior will actually be guided by those values.... Despite the evidence, the American faith in education as a cure-all persists anyway. (Robertson, 1989, pp. 278-279)

In fact, MacDonald et al. (1990) found that college students who were well informed about the prevention of AIDS continued to engage in unsafe sexual practices. Twenty-eight percent of the college sample reported having more than 10 sexual partners in the last year and even fewer reported that they regularly used condoms. In a more recent study, Hobart (1992) found that objective knowledge of AIDS and personal knowledge of AIDS victims did not

increase safe sexual practices among college students. He concluded that "those most sexually experienced know the most AIDS victims and appraise the AIDS threat most seriously, but it is also true that these experiences are inversely related to professed safe-sex practices" (p. 431).

One generalization based on these studies is that people with more schooling engage in high risk sexual behaviors, placing them at risk for sexually transmitted diseases such as AIDS. Apparently, educational attainment inadvertently promotes the very behavior that society seeks to prevent through higher education. This would mean that preventive programs should be designed to include all the population, including those with higher education.

Of interest to this particular study is how level of education affects sexually promiscuous behaviors that are associated with the transmission of HIV/AIDS. To answer this question, two studies were conducted, one at the individual level and another at the aggregate level. The first study focused on how level of education affects sexually promiscuous behaviors at the individual level. Based on these findings, a second study was conducted to examine the relationship between a region's level of education, the sexual practices of the population, and the incidence of AIDS.

Study 1

A Conceptual Model of Education and Sexual Promiscuity

The literature on sexual attitudes and behaviors not only indicates that level of education and liberal sexual attitudes are positively related but also that liberal sexual attitudes and sexual promiscuity are closely interrelated (Hobart, 1989; Hurn, 1985; Maranell, Dodder, & Mitchell, 1970; Middendorp, Brinkman, & Koomen, 1970; Reiss, 1967). Robertson (1989) concludes from a literature review on permissive attitudes that "In general, permissive attitudes correlate strongly with youth and education; older or less educated people tend to take a more conservative stand" (p. 150). Moreover, people with more liberal sexual attitudes are more likely to engage in sexually promiscuous behaviors (Allen, Onorato, & Green, 1992; Garland, Gorham, Cunnion, Miller, & Balazs, 1992; Hiatt, Capell, & Ascher, 1992; Rogers, 1992; Walter et al., 1992). MacDonald et al. (1990) found that a more casual attitude toward sex was associated with a higher number of partners among both men and women. Smith (1991) also argues that promiscuous behavior is "accompanied by increased social acceptance of premarital sex" (p. 106). Using data from an American national survey, Smith found that abstinence and to a lesser degree, lower frequency of intercourse, were associated with less education and conclude that "Respondents with less education are significantly more likely to be at no or low risk" in terms of AIDS (p. 105).

Generally speaking, people with higher levels of formal education are more likely to engage in sexually promiscuous behavior because they are less constrained by traditional norms. Those who are more educated tend to be more liberal in their attitudes and consequently are more likely to engage in nonconformist behaviors. Conversely, those who are less educated are presumed to be more conservative in their attitudes and are more likely to be constrained by traditional forms of social control. Consequently, they are less likely to engage in nonconformist behaviors.

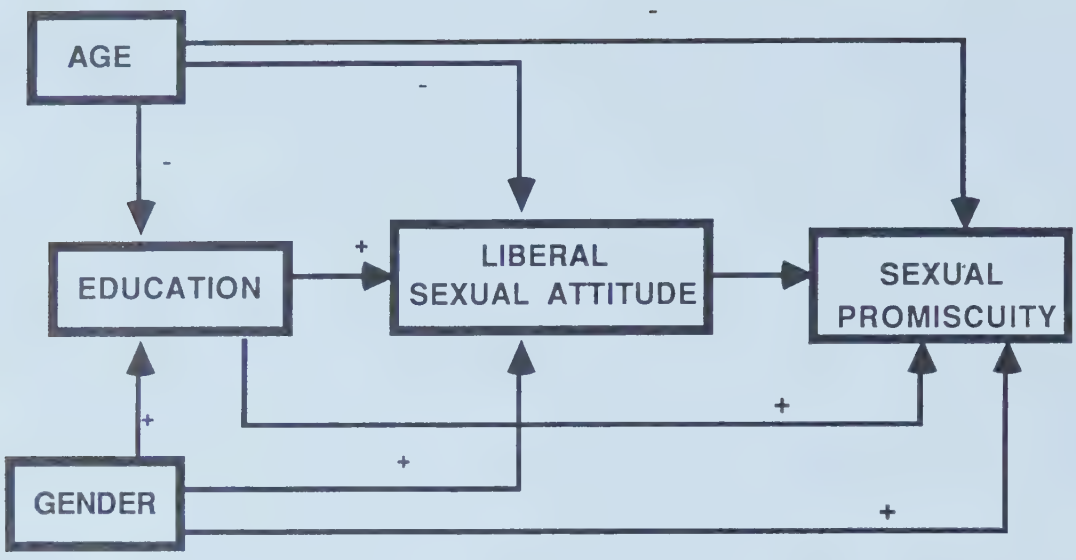


Figure 1. A conceptual model of the impact of level of education on sexually promiscuous behavior at the individual level.

In terms of the present study, level of formal education is expected to have both a direct and an indirect effect on promiscuous sexual behavior (see Figure 1). Educational attainment indirectly affects sexual promiscuity through liberal sexual attitudes. Thus it is expected that individuals with higher levels of education are more likely to have a liberal sexual attitude that increases their chances of engaging in sexually promiscuous behavior.

Method

Data

The variables analyzed in this article are taken from the 1991 General Social Survey consisting of data from a full probability sample of 1,285 adults (18 years and over) in the United States. Only those respondents (N=595) who reported that they were not married were used in the analysis. The subsample included those who were never married, separated, divorced, and widowed. The respondents were asked various demographic, attitudinal, and behavioral questions including measures of sexual behavior and attitudes.

The variables selected for analysis were level of education, liberal sexual attitude, and sexual promiscuity. Gender and age were also included in the analysis because these variables may confound the effects of the major predictors. The following age groups were used for the analysis: 20-29, 30-39, 40-49, 50-59, and 60-69. Level of education was measured as the number of years of formal schooling (0 to 20 years). The variable was collapsed into the following equal interval categories: 3-5, 6-8, 9-11, 12-14, 15-17, and 18-20 years. Liberal sexual attitude was measured by the question "Do you believe that premarital sex is always wrong (coded as 0), almost always wrong (coded as 1), wrong only sometimes (coded as 2) or not wrong at all (coded as 3)?" Four questions were selected as indicators of sexual behavior:

1. How many sexual partners have you had in the last 12 months?[1]
2. About how often did you have sex during the past 12 months? (coded as: 0= two times or less; 1=between one and three times a month; 2=once a week or more)?

3. Have your sexual partners in the last 12 months been: exclusively male; both male and female; or exclusively female? (sexual orientation) and
4. Have any of your sexual partners been casual dates or pick-ups?[2]

The data were analyzed with an R-factor analysis program in SPSSX, which is based on correlations between the four variables. It is the function of factor analysis to identity factors that are independent of one another. The factor analysis used the Varimax method with orthogonal rotation.

Results from the factor analysis, after three iterations, identified the number of sex partners and the frequency of sex as loading strongly on Factor 1. Approximately 91% of the total variance of the frequency of sex variable and approximately 89% of the total variance of the number of sexual partners variable is accounted for by Factor 1. The other two variables, sexual orientation and casual dates, did not load significantly on any one factor indicating that these two variables measured more than one theoretical dimension. For the purpose of this research, Factor 1 was used to represent sexually promiscuous behavior. The computed factor loadings were saved in an active file.

Analysis

Ordinary Least Squares (OLS) regression method was used to generate the path coefficients in the analysis using SPSSX. Path analysis is a method that allows the simple correlations among the variables to be partitioned into direct and indirect effects. A path diagram shows the causal structure of the variables.[3] An arrow from one variable to another indicates a direct path. If an arrow connects one variable to a second and another arrow connects the second variable to a third, the path is said to be indirect. In order to calculate indirect effects, the product of the direct path coefficients is used. The path coefficients are standardized and the larger the value of the coefficient, the greater the impact of that variable. Standardized coefficients may be compared in order to assess the relative effects of the variables in the model.

Results

All the predictors were correlated with sexual promiscuity at the .05 level of significance and the independent variables were not highly intercorrelated, eliminating problems of multicollinearity. As shown in Figure 2, two variables had direct effects on sexual promiscuity accounting for about 27% of the variation in the promiscuity variable. Of the two direct predictors, age had the strongest direct effect on sexual promiscuity. The negative standardized regression coefficient (Beta) of 0.38 indicates that older people were less sexually promiscuous than younger adults. Liberal sexual attitude (Beta=0.29) had the next strongest direct effect on sexual promiscuity—indicating that people who endorsed premarital sex were more sexually promiscuous. The direct effect of gender on sexual promiscuity was not significant.

The direct effect coefficient for level of education was not significant at the .05 level, indicating that education did not operate directly on sexual promiscuous behavior. Level of education, however, had a significant indirect effect on sexual promiscuity through liberal sexual attitudes (Beta=0.11). This effect, which explained approximately 10% of the variance in the dependent variable, indicates that education increased liberal sexual attitudes, which in turn increased sexually promiscuous behavior.

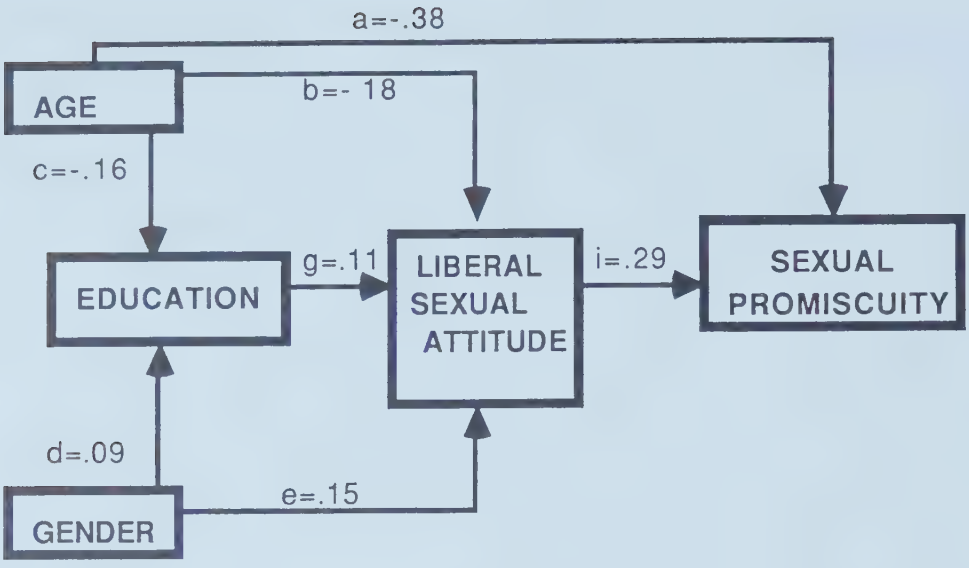


Figure 2. Observed model of the impact of level of education on sexually promiscuous behavior at the individual level.

Age and gender each had three indirect effects on sexual promiscuity. The total indirect effect of age on sexual promiscuity was –0.06 making the total effect –0.44 for age. Gender had a total indirect effect of 0.05 on sexual promiscuity.

Discussion

The results of Study 1 gave partial support to the conceptual model of educational attainment and sexual promiscuity. Level of education did not have the anticipated direct effect on sexually promiscuous behaviors. However, the results indicate a significant indirect effect of level of education on sexually promiscuous behavior. Educated people endorse a more liberal sexual attitude that is associated with sexual promiscuity in this group. Although education has a relatively small but reliable effect on individual behavior, the aggregated effects of these behavioral differences may be significant in terms of placing a population at greater risk in terms of HIV/AIDS.

Study 2

A Conceptual Model of Education and AIDS

High educational attainment in a population is a pervasive feature of the industrialized world. During the course of modernization, education increased the state of moral liberalism by promoting free enquiry and encouraging individuals to question the existing social order (Krull & Trovato, in press). However, high levels of education in a society are often obtained at the cost of tradition. In terms of sociological theory, Durkheim (1951) maintained that "Men generally have the desire for self-instruction only in so far as they are freed from the yoke of tradition; for as long as the latter governs intelligence it is all-sufficient and jealous of any rival" (p. 162). The educated person becomes increasingly independent, detached from, and less subordinate to the traditional forms of social control such as the family and the church (Trovato, 1988). Thus social control over moral behavior decreases in a society as education and

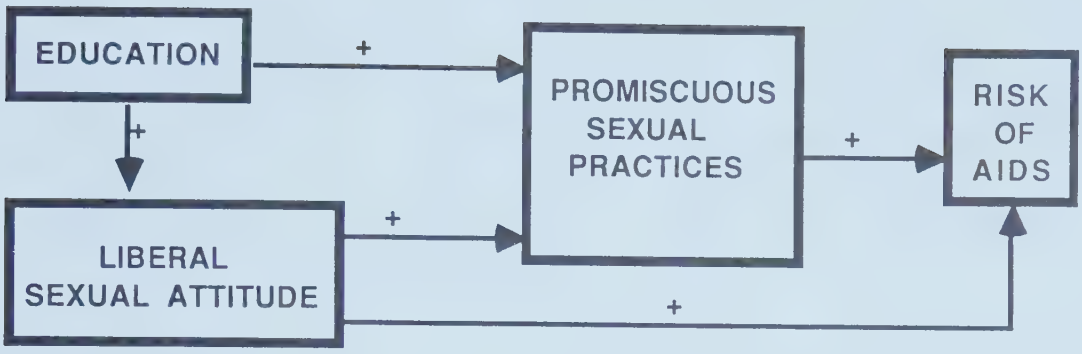


Figure 3. A conceptual model of the impact of level of education on the incidence of AIDS at the aggregate level.

liberal attitudes increase. Although education and liberalism are necessary in terms of modernization, these conditions also are associated with a greater freedom of individual expression including a wider variation in sexual practices. The sexual practices that are of interest to this study are those that are considered to be high risk behaviors associated with the spread of HIV/AIDS.

When analyzing regional variations in the rates of AIDS, it is important to look at the characteristics of the regions. The United Nations (1991) suggests that "Although a roughly exponential increase [in the rates of AIDS] is unmistakable, the rates of growth vary by region, owing to a combination of regional differences in the pace at which the disease is spreading" (p. 23). In addition to the interregional variation in the incidence of AIDS, level of education, liberal attitudes and promiscuous sexual practices vary across regions. These conditions occur in varying degrees and at different times depending on the modernization of a given region.

As shown in Figure 3, the basic argument is that regional variations in the reported cases of AIDS may be explained in part by a population's level of education and degree of liberal attitudes. These variables in turn increase the chances of high risk sexual behaviors associated with HIV/AIDS. Thus regions that have a high level of education and more liberal sexual attitudes are expected to have a more sexually promiscuous population and consequently a higher incidence of AIDS (see Figure 4).

Method

Data and Analysis

In order to analyze the variability in rates of AIDS, which is an aggregate measure, the other measures in the conceptual model were aggregated. The variables were placed in a file that computes age-gender summary measures for each of the nine regions of the United States. In this new file, each case is a region based on the aggregated gender-age specific data rather than individual measures of respondents in that region. Each gender-age-region file was then combined into one active file using the SPSSX MATCH FILES subcommand.

The data for the regional incidence of AIDS were obtained from the 1992 annual report published by the Centers for Disease Control and Prevention in Atlanta, Georgia. This report lists the number of individuals with AIDS by age and gender as well as the incidence of AIDS for each state. Because the data from the General Social Survey are classified by regions rather than by states, the AIDS data also had to be categorized by regions. The regions used were as

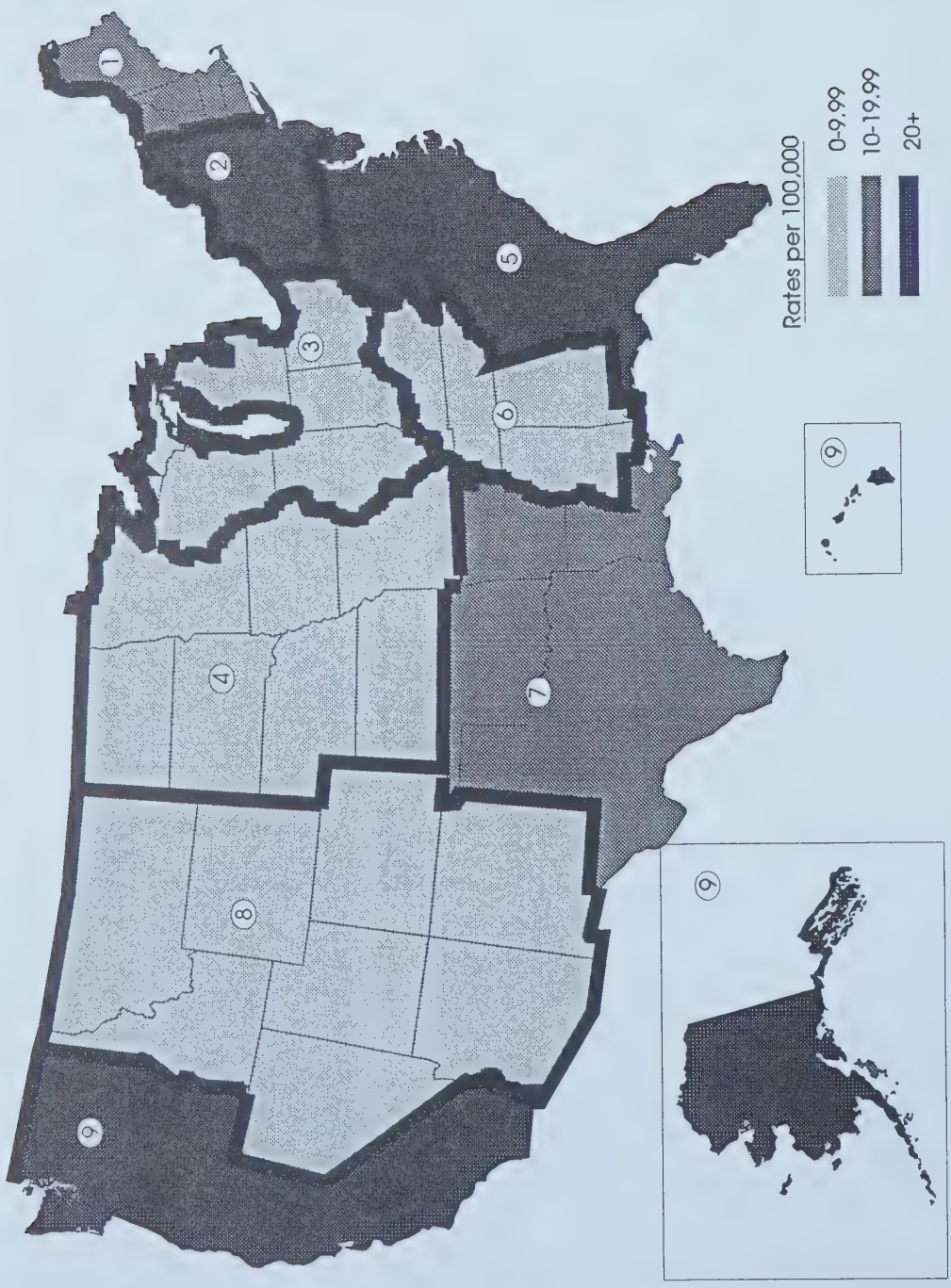


Figure 4. Regional variation in the incidence of AIDS per 100,000 population in the United States, 1991.

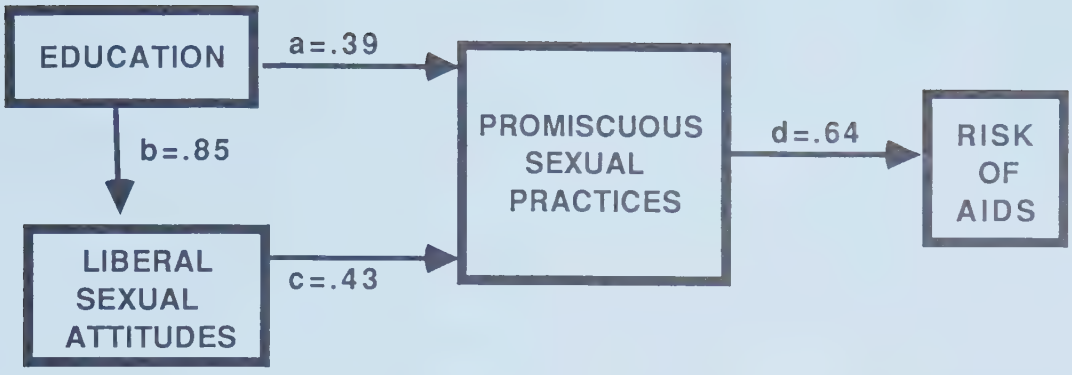


Figure 5. Observed model of the impact of level of education on the incidence of AIDS at the aggregate level.

follows: 1=New England; 2=Middle Atlantic; 3=East North Central; 4=West North Central; 5=South Atlantic; 6=East South Central; 7=West South Central; 8=Mountain; and 9=Pacific. The states included in each of the nine regions are shown in Figure 4 along with a breakdown of the regions by AIDS rates per 100,000 population.

A factor analysis was conducted on the aggregated indicators of sexual promiscuity. As with the individual level analysis, the results identified the number of sexual partners and the frequency of sex as loading strongly on factor 1. In the rotated solution, factor 1 accounted for 90% of the variance in frequency of sex and 89% of the variance in number of partners.

The data for this study conform to a cross-tabular structure making the unit of analysis a Region•Gender•Age specific cell (N=90 cells). Multivariate path analyses were then conducted on the aggregate file in accord with the variables in the conceptual model.[4] Because the unit of analysis included both age and gender, these variables were not entered as controls in the causal model as they were in Study 1. Only sexual promiscuity was allowed to directly affect the incidence of AIDS because the HIV virus can only be transmitted by sexual behavior as shown in the conceptual model. It does not make substantive sense to argue that education or liberal sexual attitudes directly cause AIDS. These variables must indirectly affect AIDS through their impact on sexual promiscuity.

Results

As shown in Figure 5, promiscuous sexual practices had a significant direct effect on AIDS (Beta=0.64) accounting for 41% of the variance in the age-gender-region specific rates of AIDS. This coefficient indicates that regions that have a high rate of promiscuous sexual practices also have a high rate of AIDS.

In terms of the effect of educational attainment on AIDS, there were two indirect paths. The first indirect effect of education on AIDS was through liberal sexual attitudes and promiscuous sexual practices (Beta=0.24). Compared with regions with lower educational attainment, regions with higher educational attainment have more liberal sexual attitudes, greater promiscuous sexual practices, and a higher incidence of AIDS. The second indirect effect of education on AIDS was through promiscuous sexual practice (Beta=0.25). Thus regions with a high level of educational attainment have a

high incidence of sexual promiscuity that is associated with a greater incidence of AIDS.

In terms of the impact of education on promiscuous sexual practices, the direct effect (Beta=0.39) and the indirect effect (Beta=0.37) are of similar magnitude and contribute to a total effect of 0.70 on promiscuous sexual practices. Level of education and liberal sexual attitudes explained about 62% of the variance in sexual promiscuity. Thus regions with higher educational attainment had more liberal sexual attitudes that resulted in greater promiscuous sexual practices than regions with lower educational attainment. Finally, level of education had a strong effect on liberal sexual attitudes (Beta=0.85) and explained approximately 72% of the variance in the attitude variable.

Discussion

The results support the conceptual model of educational attainment and AIDS. As expected, populations with higher education are more sexually promiscuous than populations with less education and consequently have a higher incidence of AIDS. Moreover, populations with a high level of education are also characterized as having a more liberal sexual orientation that coincides with more promiscuous sexual behaviors and a higher incidence of AIDS.

A discrepancy between the causal models of Studies 1 and 2 is that the direct effect of education on sexual promiscuity was not significant at the individual level, but was at the aggregate level. This difference may reflect the greater variance in the measures at the individual level, creating greater error and making it more difficult to detect a significant effect. Measurement errors are expected to be reduced when data are aggregated by age, sex, and region, and this would produce a stronger relation between educational attainment and sexually promiscuous behavior in the aggregate analysis. Thus the aggregate model may be viewed as a more adequate test of the relations explored in this study.

General Discussion

Path analysis at the individual level (Study 1) showed that level of education has a significant indirect effect on promiscuous sexual behavior through its impact on liberal sexual attitudes. Thus educated people tend to favor premarital sex more than those with less education. This attitude is associated with a greater frequency of sex with multiple partners. Overall, these findings suggest that people with higher education are at risk for behaviors involved with the transmission of HIV.

The results at the aggregate level (Study 2) indicated that the incidence of AIDS in a region was partly explained by the educational attainment, sexual orientation, and sexual practices of the population. Overall the results suggest that educational attainment does not lower the risk of AIDS, which is contrary to popular belief. Indeed, populations with high educational attainment are more at risk of AIDS than populations with less formal schooling. These findings are in accord with MacDonald et al. (1990) and Smith (1991) who reported that college students were sexually promiscuous and at some risk for AIDS.

Policy makers need to address the fact that formal education has not substantially reduced the incidence of AIDS. In fact the results from this study

suggest that people with higher education should be targeted by programs of AIDS prevention. More highly educated people hold liberal attitudes that make them likely to engage in risky sexual behavior. AIDS prevention programs in schools or through the media will need to change the sexual attitudes of well-educated populations.

To change sexual attitudes it is necessary to provide explicit instruction on sexual behavior and its relationship to AIDS. Because AIDS prevention programs are few and relatively new, most of the respondents in this study probably received little training on how to handle sexual situations that could lead to HIV infection. This absence of explicit AIDS education probably had the greatest impact on well-educated people because their higher level of schooling inadvertently encouraged general liberal attitudes. In the area of sex these attitudes increased the chances of engaging in promiscuous behavior.

It is important to recognize that greater education does not have to result in attitudes that increase behaviors associated with the transmission of HIV. If HIV/AIDS education becomes a prominent component of the school system, younger people should eventually learn sexual attitudes and behaviors that reduce their risk for AIDS. At the same time, colleges and universities can encourage on-campus programs of AIDS prevention. These could be offered as part of the diploma or degree and organized with the assistance of student associations, faculty, and administration.

Unfortunately, explicit sex education in schools and institutes of higher education continues to be a matter of debate. For example, there remains ambiguity in the schools about instruction on the use of condoms as part of a health education program. Some people still are of the opinion that instruction on condom use will encourage young people to be more sexually promiscuous. Furthermore, specific education about homosexual practices has been discouraged even though homosexuals are the highest risk group in terms of contracting AIDS. In fact, Macklin (1991) reports that "Senator Helms was able to persuade the U.S. Senate to pass an amendment to the Labor, Health and Human Resources, and Education appropriations bill that prohibits the Centers for Disease Control from using any of the funds to provide AIDS educational activities that encourage homosexuality" (p. 504).

National policy makers often take the position that abstinence should be encouraged among young people as the best way to avoid AIDS. However, this position fails to take into account that about 75% of teenagers have had at least one coital experience (Macklin, 1991, p. 504). There is almost a complete absence of sex education in grade school and this means that sexually active youngsters seldom learn the correct information and skills about reproduction and sex (compare Rogers, 1992). It also means that teenagers and adults may have difficulty relating to their sexual partners in ways that promote safe and enjoyable sex.

The only way that sexual behavior will change is if AIDS prevention is dealt with more effectively and directly by motivating behavioral change. Macklin (1991) points out that the gay community in New York and San Francisco have recently been able to reduce their incidence of sexually transmitted diseases and HIV infection. The reduction in sexually transmitted diseases and HIV depended on explicit training in terms of safer sex as well as campaigns on safe

sex backed by the approval of the community. Public figures in the gay community were used as role models to instill the motivation for behavior change and social approval by significant others supported safer sexual practices (compare Macklin, 1991).

Although changes in the educational system, such as explicit instruction on sexual behaviors, are necessary if we are to reduce the incidence of AIDS, it should be stressed that the school alone cannot be expected to bring about major behavioral changes. The results from this study, being contrary to popular belief, suggest that society has placed too much faith in education as a means of controlling the spread of AIDS. Based on Hurn's (1985) review of education and social change, Robertson (1989) went on to argue that

> the faith that the schools alone can bring about social change is distinctively American. As sociologists of education are increasingly pointing out, there seems to be very little empirical justification for this faith, and it may be that we have cherished expectations of the institution that it cannot fulfil alone. (p. 279)

The results obtained here tend to support the argument that high levels of education have not protected people from AIDS. Rather, formal education has inadvertently increased those behaviors that are associated with AIDS. Incorporating changes into the educational system may help to reverse this trend, but the local community and other sectors of society must also take some responsibility. National leaders need to develop a well-integrated plan to deal with the spread of AIDS that involves all levels of society (Rogers, 1992). This plan would bring together the local community leaders, health workers and professionals, organizations such as the Centers for Disease Control, and other relevant sectors of society. The objective would be to develop government policies that would in turn regulate safer sexual practices, reduce transmission of HIV, and reduce the incidence of AIDS in the population.

At the individual level, the plan would include specific grade school instruction on safer sex: targeting youngsters who are becoming sexually active. In addition, sexual skills programs would be an important new strategy. These programs would teach young men and women how to handle awkward situations involving sex (i.e., using condoms as part the foreplay to sexual intercourse). Finally, the plan would include social modeling and incentives systems to motivate and maintain behavior change (compare Kazdin, 1989).

Future research should consider using Census data for the age, gender and education variables rather than relying on survey data. It would also be of interest to include variables such as race, income and population density in the model. Given the increasing incidence in reported AIDS cases in Canada, it would be worthwhile to conduct a similar study using Canadian data. Although Canada has reported a lower incidence of AIDS in comparison to the United States, the educational policies recommended in this article should be considered in the Canadian context.

Acknowledgments

This research was supported by a fellowship from the Social Sciences and Humanities Research Council of Canada. I wish to thank Dr. P. Krishnan and Dr. W. David Pierce, Department of Sociology, University of Alberta, for their helpful comments. I would also like to thank Dave Odynak, Population Research Lab, University of Alberta, for his help in producing the graphics.

Correspondence should be sent to the author, Department of Sociology, University of Alberta, T6G 2H4.

Notes

1. Two different formats were used to obtain information on this question. The first was open-ended and the second closed-ended. Some respondents answered the closed-ended question and others answered the open-ended question. The open-ended question was coded to match the categories of the closed-ended question and then the data obtained for the two questions were combined.
2. Indicators of religiosity were also used in the original factor analysis but were excluded from the regression analysis for this particular study in order to focus more specifically on the effects of education.
3. Some researchers may prefer an analysis of the data based on LISREL. One difference between LISREL and path analysis is that LISREL searches for causal relationships whereas path analysis already assumes causal ordering. Because a weak causal order among the variables can at least be assumed here, path analysis is an appropriate statistical procedure. In addition to the assumption of weak causal ordering, the researcher must also assume that the relationships among the variables are causally closed, and this is tenable in the present analysis. Path analysis is primarily a method of working out the logical consequences of the two foregoing assumptions (Kim & Kohout, 1975, pp. 383, 385). The fit of the path model can be assessed from the R2 coefficient of each model. The R2 for a model may be interpreted as the proportionate reduction in error in predicting the dependent variable from knowledge of the independent variables operating jointly.
4. Survey research generates a cross-sectional data structure. Because of this limitation, there is always some concern about the time ordering of the variables. A researcher cannot be sure that the predictors actually preceded the variation in the dependent variable. Also, the aggregation of survey data is sometimes problematic because regions may not be well represented when individuals are the sampling units. Census data are preferred for aggregate analysis, but attitudinal and behavioral measures are usually not included in the census information. The present study is conducted within these limitations.

References

Allen, D.M., Onorato, I.M., & Green, T.A. (1992). HIV infection in intravenous drug users entering drug treatment, United States, 1988 to 1989. *American Journal of Public Health, 82,* 541-546.

Burke, D.S. (1990). Human immunodeficiency virus infections in teenagers: Seroprevalence among applications for US military service. *Journal of the American Medical Association, 263,* 2074.

Centers for Disease Control. (1991). *HIV/AIDS surveillance report.* Atlanta, GA: Author.

Centers for Disease Control. (1992). *HIV/AIDS surveillance report.* Atlanta, GA: Author.

Centers for Disease Control. (1993). *HIV/AIDS surveillance report.* Atlanta, GA: Author.

Durkheim, E. (1951). *Suicide: a study in sociology.* J.A. Spaulding (Trans.), with an Introduction by G. Simpson (Ed.). New York: Free Press.

Garland, F.D., Gorham, E.D., Cunnion, S.O., Miller, M.R., & Balazs, L.L. (1992). Decline in human immunodeficiency virus seropositivity and seroconversion in US navy enlisted personnel: 1986 to 1989. *American Journal of Public Health, 82,* 581-583.

Hiatt, R.A., Capell, F.J., & Ascher, M.S. (1992). Seroprevalence of HIV-type 1 in a Northern California health plan population: An unlinked survey. *American Journal of Public Health, 82,* 564-566.

Hobart, C. (1989). Premarital sexuality. In G.N. Ramu (Ed.), *Marriage and the family in Canada today* (pp. 53-75). Scarborough, ON: Prentice-Hall.

Hobart, C. (1992). How they handle it: Young Canadians, sex, and AIDS. *Youth and Society, 23,* 411-433.

Hurn, C. (1985). *The limits and possibilities of schooling: An introduction to the sociology of education* (2nd ed.). Boston: Allyn and Bacon.

Kazdin, A.E. (1989). Behavior modification in applied settings (4th ed.). California: Brooks/Cole.

Kim, J., & Kohout, F.J. (1975) Special topics in general linear models. In N.H. Nie, C.H. Hall, J.G. Jenkins, K. Steinbrenner, & D.H. Bent (Eds.), *Statistical package for the social sciences* (pp. 383-385). New York: McGraw-Hill.

Krull, C., & Trovato, F. (in press). Modernization, social integration and the sex differential in Quebec's suicide rates: 1931-1986. *Social Forces*.

MacDonald, N.E., Wells, G.A., Fisher, W.A., Warren, W.K., King, M.A., Doherty, J.A., & Bowie, W.R. (1990). High-risk STD/HIV behavior among college students. *Journal of the American Medical Association, 263*, 3155-3159.

Macklin, E.D. (1991). AIDS: Implications for families. In J.E. Veever (Ed.), *Continuity and change in marriage and family* (pp 496-509). Toronto: Holt, Rinehart and Winston.

Maranell, G.M., Dodder, R.A., & Mitchell, D.F. (1970). Social class and premarital permissiveness: A subsequent test. *Journal of Marriage and the Family, 32*, 85-88.

Middendorp, C.P., Brinkman, W., & Koomen, W. (1970). Determinants of premarital sexual permissiveness: A secondary analysis. *Journal of Marriage and the Family, 32*, 369-379.

Reiss, I. (1967). *The social context of premarital permissiveness*. New York: Holt, Rinehart and Winston.

Robertson, I. (1989). *Society*. New York: Worth.

Rogers, D.E. (1992). Report card on our national response to the AIDS epidemic—Some A's, too many D's. *American Journal of Public Health, 82*, 522-524.

Smith, T.W. (1991). Adult sexual behavior in 1989: Number of partners, frequency of intercourse and risk of AIDS. *Family Planning Perspectives, 23*, 102-107.

Trovato, F. (1988). A longitudinal analysis of divorce and suicide in Canada. *Journal of Marriage and the Family, 49*, 193-203.

United Nations/World Health Organization. (1991). The AIDS epidemic and its demographic consequences. *Proceedings of the United Nations/World Health Organization Workshop on modelling the demographic impact of the AIDS epidemic in pattern 11 countries*. New York: Author, Department of International Economic and Social Affairs.

Walter, H.J., Vaughan, R.D., Gladis, M.M., Ragin, D.F., Kasen, S., & Cohall, A.T. (1992). Factors associated with AIDS risk behaviors among high school students in an AIDS epicenter. *American Journal of Public Health, 82*, 528-532.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 21-34*

Keith D. Walker
University of Saskatchewan

Notions of "Ethical" Among Senior Educational Leaders

This article describes a study of the ethical definitions characterizing the moral wrestlings of members of the Saskatchewan League of Educational Administrators, Directors and Superintendents (LEADS). LEADS is a professional public service organization empowered by provincial statute to register and monitor senior educational administrators who in turn are charged with the responsibility for delivering education (elementary and secondary level) to the public of one of Canada's 10 provinces. The leaders' a priori definitions of "ethical" are described and analyzed in the research presented in this article. The article describes the ethical consciousness of senior public administrators and identifies the various meanings associated with the notion of ethical. These discrete definitional understandings provide the basis for a description of the notion of ethical and demonstrate an alternative research approach to describing the ethical orientations of educational leaders.

Cet article décrit une étude des définitions de l'éthique qui caractérise les dilèmes moraux des membres du Saskatchewan League of Educational Administrators, Directors and Superintendents (LEADS). LEADS est une organisation professionnelle de services publiques sanctionnée par statut provincial pour indiquer et surveiller les administrateurs éducatifs séniors qui, à leur tour, sont chargés de la responsabilité de l'éducation (des niveaux élémentaire et secondaire) envers le public dans une des dix provinces du Canada. Dans cette recherche, les définitions a priori de "l'éthique" d'après ces leaders en éducation sont données et analysées. L'article décrit la conscience éthique de ces administrateurs publiques séniors et identifie différentes significations associées à cette notion d'éthique. C'est en se servant de ces définitions discrètes et personnelles que l'on arrive à formuler une description de la notion de l'éthique et qu'on puisse démontrer une méthode de recherche alternative pour décrire l'orientation de ces leaders en éducation qui serait conforme à l'éthique telle que décrite.

The purpose of this article is to offer some insight into the meaning of everyday ethical decision making from the perspective of senior educational leaders. The guiding parameter for this research was related to exploring the meanings associated with the notion of ethical by educational leaders. The research methods included the use of surveys and intensive interviews.

The field of educational administration has only rarely researched the nature of ethical decision making in an exploratory and descriptive fashion. This is to say, with Gronn (1987) and Enns (1981), that there have been few examples of "systematic attempts to describe, analyze and explain everyday administration as it [is] experienced by people, let alone to seek to learn from those experiences" (Gronn, 1987, pp. 105-106) and that past "approaches for understanding and shaping human activities ... have failed to take full account of the major dimensions of human existence, namely, the ethical-moral" (Enns,

Keith Walker is with the Department of Educational Administration. His teaching and research interests concern applied and professional ethics, the philosophies and politics of educational leadership, and pastoral care in educational communities.

1981, pp. 1-8). Certainly their assessment requires some significant updating as one thinks of Sergiovanni (1990, 1992), Leithwood and Musella (1991), Hodgkinson (1991), Maxy (1991), Ashbaugh and Kasten (1991), Smyth (1989), and Foster (1986) who have recently begun to address (or readdress) the ethical domain in educational administration. These more recent educational leadership books seem to be doing more than "flirting with values" (Macpherson, 1985, p. 20). Leithwood and Musella (1991) were able to locate only 38 studies on the superintendency. These authors have noted that previous superintendency researchers have given their greatest attention to leader practices, with much less attention to internal processes (attitudes, values, and beliefs). This study addresses this deficit by considering the specific nature of ethical decision making of superintendents (N=188), as understood from their definitional perspectives.

The usual research approaches in applied ethics have been oriented toward observation of behaviors or the analysis of participant responses to preconceived situations or ethical challenges. This article complements such orientations through a methodology that allowed the leaders to voice their own meanings and understanding with respect to the nature of ethical decision making. Toffler (1986), Bird and Waters (1987) and Jackall (1988) are representative of the very few researchers who have sought to develop first-hand reports of the nature of ethical decision making among administrators. This article describes a research project that set out to let administrators speak for themselves rather than asking them to measure up to a priori ethical templates.

The survey respondents were directors (44%), superintendents (21%), assistant superintendents (20%), and regional directors of education (5%). With respect to the interviewees, more than half were from rural divisions and represented divisions ranging from 10 schools to over 60 schools in size. In total the 20 directors had 5,161 instructional staff and 86,513 students in their jurisdictions. The leaders were asked to supply their views on the definition of ethical. Two interviews were conducted with each of 20 randomly selected directors of education to discover, explore, elaborate, and interpret their definitions. Further data were collected from the organization's archives.

I attempted to go beyond superficial descriptions (*begreifen* in Ladd, 1957) to look, rather, at leaders' understandings from an inside perspective (*verstehen*). As Toffler (1986) explains in the introduction to her study on the ethics of managers, "the absence of ... first hand reports—and the need for them in order to expand both knowledge ... and ... practice itself" (p. 3) prompted the development of her study. This presentation reports findings that respond to the question: "what do educational leaders say ethical decision making in their profession means?" Bird and Waters (1987) have alerted those researching in the ethical domain that leaders will not likely be systematic or traditional in their use of ethical language. This article organizes the data collected from these leaders into categories and takes due care to guard the integrity of the meanings and contexts of their particular ethical wrestling. Jackall (1988), in his attempt to report on the first-hand moral experiences of corporate managers, says, "managers may disagree with some of the broader interpretations of their experiences suggested here. I have tried, however, to capture the complexities, ambiguities and anxieties of their world" (p. vii).

Similarly, I attempt to provide the reader with a sense of the definitional nature of ethical decision making among the members of the Saskatchewan League of Educational Administrators, Directors and Superintendents (LEADS). It is interesting to note that the LEADS organization derives its responsibilities from its own provincial statute and code of ethics. The League is legally and morally mandated to ensure, among other things, that high ethical standards are sustained among its members.

Definitions for the Notion of "Ethical"

The question addressed by this article asks: how do educational leaders define the word *ethical*? Although the meaning of such a term of reference is often assumed, the article suggests that there was indeed a variance in the understandings associated with this common descriptor. I used both direct and indirect means for deriving definitions from the sample.

Direct Definitions from Survey

In response to a survey question that asked, "according to your use and understanding, what is the meaning of the word *ethical*?" the responses ranged from objective and external types of references to subjective and personalized types of references. Some respondents expressed definitions related to philosophic or principled truth, whereas others oriented their responses to thinking about the word *ethical* in terms of specific rules of behavior. Some respondents inclined their definitions to a more generic dictionary-type definition. In other words, some respondents considered the source (epistemological), some the reality (ontological), and some the definitional (synonymic) character of the word *ethical*.

Ethical as subjectively or objectively derived. It was most common for leaders to describe ethical as either objectively or subjectively derived. By the term *objective* we mean that the responses referred to some code, principle, or point of reference beyond or external to the individual decision maker. For example, some respondents said that ethical was: "conforming to the standards of conduct of a given profession," "acting in a fashion that takes into account the standards set legally within the Education Act and also with the [professional organization's] code of ethics," "based on Christian principles" or "moral reasoning and behavior that reflect 'universally' accepted principles of conduct." Those responding with these types of definitions seemed to tie the source or origin of standards or expectations to objects beyond their own person. In other words, they viewed the concept of ethical in terms of principles, guidelines, or codes that they had derived or discovered from societal, professional, or religious sources. This view held that the meaning of ethical was independent of the educational leader.

Those responding with subjectively oriented definitions of ethical expressed a leader-dependent understanding of what was meant and what was not meant by the term *ethical*. For example, some respondents said: "I behave ethically when I do things that make life better for me and people around me," "[ethical is] behavior harmonious with the ideals of a democratic society that include: sensitivity to right and wrong; the use of one's personal philosophy and beliefs as a guide to individual behavior; and a regard for the dignity and worth of every human being" or "[ethical is] doing what is right ... listening to

your conscience." For those responding along these lines, the emphasis seemed to be related to a personal, individual, or subjective sense of what might be right or wrong, good or bad, virtuous or vicious.

Some responses combined these objective and subjective orientations. For example, some respondents said that the word *ethical* refers to: "behavior and actions which are made on the basis of certain beliefs and principles held by an individual. These beliefs and principles in our society are consistent with the Golden Rule and the principles of natural justice," "those 'rules' or practices (written or unwritten) which guide an individual in both personal and work life" or "behavior which is congruous with our own personal, or more importantly, a professional code of values or ethics." These respondents indicated that both personal and extrapersonal ways and sources of knowing provide the criteria for determining what might be appropriately labeled "ethical." One might infer from these definitions that, for these respondents, ethical behavior consisted of "harmonizing" or blending the two (internal and external) ways or sources of understanding. Forty-two percent of respondents chose to express an objective definition, 32% indicated a subjective definition, and 23% of the educational leaders replied with a conjunctive definition.

Ethical as rational or behavioral. Two additional ways of categorizing responses to the direct request for a definition of the word *ethical* are related to rational and behavioral types of responses. Those definitions oriented to principles, processes or the person's state of being are categorized as rational type responses. Another type of response emphasized observable behavior or its effects. Some definitions contained elements of both the rational and behavioral orientations.

In the former instances the respondents typically provided a more general definition of ethical and offered some virtues or principles that they equated with the status of "being ethical." For example, ethical was said to be: "consistent with being morally right; without ill-malice and based on sound principles," "to make decisions that are fair and honest—treating other people with dignity and respect" or "it means responsibility, fairness, justice, compassion and rights."

On the other hand, behavior and conduct-oriented responses such as the following were also expressed: "acting in a morally responsible fashion" or "to act within the parameters of acceptable and standard guidelines." The reader will notice that these definitions focus on the educational leader's conformity to sets of behavioral rules or codes of conduct, whereas in the case of rational definitions the emphasis was on ideational constructions.

The subtle differences between these two set of definitions might be simply expressed as the difference between general ethical aspirations and more specific ethical actions. In some instances the responses contained elements of both behavioral expectation and rational imperatives. For example, respondents said: "ethical refers to behavior and actions which are made on the basis of certain beliefs and principles held by an individual. These beliefs and principles in our society are consistent with the Golden Rule and the principles of natural justice" or "being ethical means that one abides by a moral code." It is to be noted that the above definitions sought to harmonize thinking and acting through choosing.

It may be said that survey respondents usually expressed their definitions of the word *ethical* in either rational or behavioral terms. Fifty percent of leaders considered ethical to be related to rules of conduct or behavior and 38% indicated that they viewed ethical in rational or philosophic terms, having to do with the state of being of persons rather that the state of doing. The survey results indicated that 10% of leaders held to a definition that had elements of both orientations.

Ethical as analogous. The survey respondents also seemed to define the word *ethical* by simply supplying equivalent phrases, much like one would find in a dictionary. These responses indicated the respondents held that the word *ethical* was clearly related to describing the rightness or wrongness of an act or attitude. Twenty-three percent of responses were so stated, as indicated by the following examples: "the just, fair, and appropriate thing to do. Moral behavior," "morally correct behavior in a professional setting," "making the 'correct' decision, based on high standards of moral reasoning" or "ethical is what is morally correct—in simple terms ethical is to do what is right."

Direct Definitions from Interviewees

When those being interviewed were asked to define what the word *ethical* meant to them, their responses provided some further elaborations to those described through the survey responses. Those interviewed related the term *ethical* to one or more of the following themes: community values, dichotomies, core ethical values, and the difficulties inherent in or resulting from the concept of ethical.

Ethical as community values. One community-values conscious director said,

> I think for us, the people elected to be on the Board represent the community. They bring the community ethic with them so far as what is "right" and what is "wrong" in that community. Some years ago, if a teacher's car was parked in front of the pub that was a wrong thing—there are those kinds of things [that define ethics]. They expect their employees to live within ethical boundaries. They don't get into specifics like marriage is prerequisite, if living with someone, but if you go into the community as a staff member and you stir up something that really is going against the direction that the Board wants the school to go—they aren't happy.

This director defined ethics in terms of the community mores or values. Ethical boundaries were not specifically stated but taken for granted, until they were transgressed. Even after a transgression of community values had been identified, the actual ethical criteria may not have been explicitly stated with the various forms of censure or discipline. From an educational perspective, the director held that the board defined these ethical boundaries and that these had changed over the years. At the time of these interviews, the director felt that the community ethic had become less precisely defined. It would seem that stirring people up and disrupting the community might produce a definition of specific "rights" and "wrongs." The director believed he should live within the community boundaries, with other employees, in order to be consistent with the board of education qua community. Ethical, then, was conceived of as a set of boundaries established by the community. The reader will note that this view of ethical is effect-oriented. Ethical consideration comes into play

when the status quo is upset or, in other words, when a boundary is transgressed. Other directors spoke of the notion of ethical from the perspective of the school system and the larger community. One used the words "kid-centered, rural, small 'c' conservative, family-oriented and ten commandment-types" to provide a definition for the term *ethical*.

Ethical as conceptual dichotomies. A number of the directors indicated that they felt there were some concept dichotomies that helped them to define the term *ethical*. These dichotomies tended to spell out the limits of the term ethical both contextually and conceptually. For example, certain directors viewed ethical values as a rather wide classification of values that could be held by anyone, whereas morals were "more or less Catholic," in the church sense. "While we constantly live with the ethical, when you are working with people," this leader thought, "the ethical realm is more obvious." This administrator suggested that if one was working with people, one was obliged, ethically, to treat them "fairly and correctly." Ethical was, then, defined by the right or wrong way of dealing with people or issues. The leader said, "I think morals are more or less Catholic values whereas ethics are anyone's values—it's just what is right and wrong. I think you live in the ethical realm all of the time. It is more obvious in some contexts. For example, when dealing with something people-oriented—you've got to treat people fairly and correctly."

Another leader also differentiated between morals and ethics but suggested that for him morality "has to do with one's life-decisions" and ethics has "to do with honesty ... public good, unselfishness and professional behavior." Although he understood ethics as the wider term, his experiences and the results of past behavior had informed his ethical beliefs such that what he considered to be ethical had "become tighter and tighter." In other words, this director felt that he had become less tolerant of some behaviors. He felt he could more easily make definite decisions that he had considered difficult to make at points earlier in his career. On the other hand, moral decisions—having to do with "good or bad, sin versus doing right"—had become fuzzier. He was more hesitant about making moral decisions than he had been earlier in his life. Ethical would appear to have been related more to the public or social life and career of this leader, whereas the morality of the leader was related to personal or private behaviors or decisions. He said,

> I see a difference between morals and ethics. Morals are related to good or bad, sin versus doing right ... It has to do with one's life decisions. Ethics I consider having to do with honesty and with the public good, unselfishness and professional behavior. I see ethics as a larger area than morals and I used to see morals as black and white issues—but now I see the whole area of morality as a less definite area, whereas, in terms of ethics, my ethical beliefs and what I tolerate as ethical has become tighter and tighter during my career. This is partly because of experiences, partly because I see that some behavior has negative results and some has positive results so I am able to make definite decisions about ethical behavior that I couldn't make 20 years ago ... I am unable to make decisions about moral behavior that I wouldn't have hesitated to make 20 years ago.

Ethical as common life principles. Other directors expressed their view concerning the meaning of the word *ethical* by indicating that certain core principles were ethical in nature. These leaders were consistent in their view that

one must both ascribe to these core life principles and practice them to earn the descriptor *ethical* as a characterization of one's leadership. One director viewed ethical as "embedded" in the ways that people in the organization thought and behaved in relation to others. His organization had at his initiative developed a policy that expressed a three-component definition of the duties associated with being ethical. These he called "duties of care" or "caring principles." Due regard for individuals, professional practice, and due process constituted this ethic of care. He expressed himself in the following manner:

> I would think the ethical issues are embedded inside other things. I tend to try to think of ethical concerns as fundamental human issues. We have a policy developed as a part of the philosophy statement called the "duty of care." In it, we suggest that everyone in the organization has the right to care. The three components to the duty of care are due regard for the human individual, diligence and hard work as one carries out their professional obligations and, thirdly, due process or fair treatment. We argue that these [ethical] principles should underlie everything we do.

Several directors saw ethical as being related to fair treatment of people and the avoidance of harmful effects in decision making. Others suggested, similarly, that to be ethical one must be "fair and square" as well as being perceived to be so. One leader equated ethical with the notion of integrity. His view was very close to that of the previously quoted director who also expressed his belief that "systems [are] all connected" and therefore one should appreciate that each participant in the system has the potential of affecting the whole. The integrity-oriented leader believed that one must not be seen as self-serving, but rather be oriented to the good of the students and the staff. Ethical, then, was related to the leader's capacity to actually be, and to be perceived to be, one who would put the good of students under care as his or her first priority. This leader said, "integrity is the first of all moral considerations. One must be perceived to be for the good of the students and staff. When one becomes self-serving then the leader begins to lack integrity."

Ethical as difficult. The final theme emerging from the direct question "what does the term *ethical* mean to you?" pointed to difficulties. For some the descriptor *ethical* applied to problems that were difficult. One interviewee, for example, understood ethical to be related to situational conflicts, dilemmas, or grey areas wherein "there are no easy answers." He referred to being faced with "the ethical" when being "caught" with no previous experience or guidelines and having to work through the problem by "building on hunches." Ethical problems were, to some, understood to be value judgments emerging from situations that are not already resolved or settled.

One of the interviewees defined the notion of ethical as the source of some consternation for both him and his board because of differences in their respective views. His board saw issues in a more black and white fashion than did he. The leader himself saw ethics more as dealing with noneducational issues such as abortion, premarital sex, and cheating on income tax returns than with educational problems. Ethical was thought of as being a matter of right or wrong and demanding a weighing of arguments to determine permissibility. He had difficulty with seeing educational issues as clearly ethical in this sense. Although he acknowledged that such issues were probably in the education

system, he believed that his own involvement in the system made educational sorts of issues rather "grey-oriented." His board members, on the other hand, were more definitive in their identification of the notion of ethical as related to issues and dilemmas in the educational realm.

Indirect Definitions of Ethical from Interviewees

Ethical as heroes and villains. The interviewer asked each director of education two questions concerning their view of what might constitute the meaning of the notion of ethical. These two questions concerned describing ethical heroes and ethical villains. The interviewee responses provided me with data from which the characteristics of an ethical and an unethical person might be indirectly derived. During the interviews directors were asked what it would take to be an ethical villain in their school jurisdiction. In other words, what would I have to do to be considered an ethically bad person? This question was oriented to determining the directors' view of what the organization would consider to be sacred and what might be considered to be profane. In some instances the respondents referred to particular behaviors, some to attitudes and affiliations, some to attributes, some to attentions and intentions. Each of these descriptions provided an indirect insight into these leaders' views of what was meant by the words *ethical* and *unethical*. It is interesting to note that the directors of education could much more readily describe ethical villains than heroes: this, then, provided a thicker negative description of the concept than it did a positive one.

One director spoke with emotion as he reflected on a person whom he considered to be an outstanding example of an ethical hero. This person loved and cared for children. He was a naturally effective school leader who worked hard and made relationships a priority. This man was ethical because he recognized the importance of people, especially schoolchildren, and worked exceptionally well with them. In the director's view the principal had sacrificed some personal and professional goods to the benefit of "the kids":

> I can think of an ethical hero in our division. He finds a way of embracing every kid who comes into his school—he loves them and cares for them. He's never read any of the effective school literature but he does everything. He's got relationships together and is a real hard worker. He could do my job (director) better than me—but he chooses to do what he's doing.

Some ethical villains were defined by their exploitation of weaker persons for purposes of self-gratification as in the instance of sexual abuse. Some ethical transgressions were deemed so because certain persons had extended their roles beyond appropriate limits. These ethical villains ranged, then, from invaders of persons (as in cases of child abusers) to invaders or interlopers of role responsibilities (as in board members "doing administrative mischief"). To provide the reader with other examples of the perceived range of behaviors and attitudes considered commensurate with the label "ethical villain" or "ethical hero" several responses are displayed below:

> In our division we probably have one major ethical villain a year that is dealt with. Deliberate lies, falsifying reports and major theft are considered ethical infringements by our Board. The abuse of children is the worst of all transgressions. Buying booze for kids is an example of a past ethical infringement that

> resulted in dismissal. Not following explicit directions of Board or Director is considered [ethically] wrong.

Another educational leader said, "making out with a high school girl would put you on the road pretty quick ... Screwing around with the finances ... Misleading on your expense claims. Undermining the Board in a deliberate way."

Ethical as mandate-oriented. More specific "laws of ethics" were formulated by the directors related to the concepts of a mandated duty or public trust that required more of an educational leader than mere legal compliance. These notions were generally centered around the leader's responsibility to be child-focused. In instances where this mandate-oriented ethic was presented as an aid to define the word *ethical*, it was common for directors to see themselves as the guardians of the mandate and view others as more likely candidates for related "ethical failings." Failing to do one's job of providing for children (i.e., because of alcoholism problems) was cited as unethical. "One should be more responsible to the public and professional trust," according to one interviewee. He said, "anytime one's own private behavior impinges, negatively, on the welfare of children then the person should be responsible enough to censure themselves." To fail ethically had two aspects: getting into a position of not being able to do work with "kids" properly and, second, to fail to recognize and correct such a situation. One administrator defined being ethical as "doing an unpleasant duty well and professionally, in spite of personal feelings."

Ethical as people-favoring. For some educators ethical people were exemplified as those committed to the other persons in the organization. One leader said, "you become a hero here by virtue of hard work, personality and how you treat people." This administrator associated ethical with choosing to work hard and treat people in an appropriate fashion. These would appear at first consideration to be matters of choice, whereas his other feature is personality, which may be understood to be less volitionally oriented. When asked to elaborate, the director indicated that being ethical had to do with balancing work accomplishment and people advocacy. The leader's personality mediated, in his view, between these two demands, and the proper balance of these two criteria determined whether one was deemed to be ethical or not.

Ethical as a vice-avoiding. Mistreating or exploiting people, treating them with indifference, or other forms of abuse were all cited as examples of behaviors or attitudes that formed negative meanings of the term *ethical*. If one were to undermine people because of personal preferences or to try to oust someone from her or his job for insufficient cause, these behaviors would in some directors' estimations be unethical. They cited the creation of imbalances in accounts of information through false reporting or withholding important information as ethically wrong behaviors. The idea of confidentiality was often used by administrators to describe the tensions between the determination of what constitutes ethical and the unethical. The whole realm of confidentiality was considered as ethically embued. Sharing information of privilege was also seen as a transgression of professional privilege and a sign of indiscriminate betrayal of trust. One director called this "telling tales out of school." One director said, "taking comments about colleagues into the community or dis-

cussing children's behavior and abilities outside professional channels or evaluating superordinates anywhere with anyone" as examples of unethical practice. "Ethical villains are people who break confidentialities, spreading gossip," indicated one director.

Ethical as befitting a community. Many of the leaders related the notion of ethical to community values and standards. A number of directors made reference to religious beliefs. Ethical heroes were insiders and ethical villains were outsiders. An unethical person is one who practices a lifestyle inconsistent with the "basic ethical code" developed by the people [of the religiously oriented community] over the years. It was felt that those who subjected themselves, in compliance, to these community-wide belief sets would be deemed to be ethical. For example, they said, "[in our community] an ethical hero is basically somebody with the same religious affiliation, not necessarily the same church but basically the same beliefs" or "from a lifestyle standpoint a person jeopardizes their ethical status and employment by obviously defying the Church."

Other leaders conceived of the ethical hero as a consistent, task-oriented protector of the educational rather than a religious community. For these leaders the ethical person was one who could determine educational values in a consistent fashion and stick with these value-driven priorities. These directors also indicated a belief that an ethical hero appreciates the vulnerability of educational leaders and expresses a loyalty to them by immediately defending them from those who would be critical. An example of this type of community perspective is reflected in the following statement:

> Ethical heroes set priorities and goals then work single-mindedly in pursuit of these. They don't let things impinge upon these goals. They protect their fellow workers. Ethical heroes jump to the defense of their leaders who are being shot at and are consistent in their characteristic responses to decisions and situations.

Ethical as rational, but not self-interested. In contrast to the rather altruistic and community-minded definitions and descriptions above, some leaders pointed to the profane propensity of some leaders to act out of self-interest. These attitudes and behaviors were viewed with particular disdain as unethical. Examples of this self-interest were exemplified by self-aggrandizement, crossing normative role expectations, position abuse, and insensitivity to established community practices. For example, they said it was unethical "to do something because of self-interest, because you are going to gain from it and to sacrifice students." "These are people who use their position for personal gain or who present themselves as having competencies that they really don't have," said another director of education.

A number of directors expressed the view that thinking characterized by unreasonableness or thoughtlessness was to be considered unethical and a serious affront to the group's professional ethics. Further, some directors believed that unethical acts were those that would be considered to be unreasoned by the average person. An example of these meanings for ethical was related by a director who felt that the mere fact that a person would choose to rationally address issues and struggle with conflicting ethical demands constituted an act of ethical heroism. An ethical person is one who is seen by others to be thoughtful, reasonable, and willing to contend with moral challenges.

Beyond the particular choices or the moral processes that lead to "ethical" decisions, this leader felt that a leader's willingness to participate in difficult issues distinguished that person as ethical. "People," he said, "who wrestle with tough choices and demands are ethical heroes."

Summary

The leaders in this study defined the word *ethical* in a variety of ways. The first of these focused on the respondents' perception of their source of "ethical" authority, either objective (external) or subjective (internal). Definitions were also categorized by leaders as referring to either behavioral or rational constructs. Some respondents utilized synonymous terms to define their view of the word *ethical*. Definitions of ethical were often related as the reflection and alignment of community values and boundaries to common life principles and to conceptual dichotomies. The indirectly derived definitions of the word *ethical* emerged through what in leaders' estimations would constitute an ethical hero or a villain (the unethical person). Accordingly, the ethical person was considered to be law-abiding, child- and education-oriented, people-friendly, and virtuous in regard to others, a conformer to community mores and a rational-altruist. Each of these orientations provides a slightly distinctive color or texture to the meaning of the word *ethical* for the leaders who expressed their understandings.

Implications of this Study

> Ethical reflection take[s] place in the actual situations of choice and action ... Academic ethics must take these situations of human moral wrestling as the primary material of ethics and bring resources that enhance the quality of judgement and action. (McCoy, 1985, p. 22)

Implications for Theory

The implications for theory that may be drawn from these research findings are numerous. Educational leaders have described their various meanings for the term *ethical* and these reflect the diverse moral sense of educational leaders. The finding that leaders differ in their understandings of what constitutes ethical is not to be interpreted to mean that these various meanings are in conflict. The more specific nature and relationship of the various notions of ethical warrant further investigation. For example, what are the presuppositional grounds that accompany these different understandings and expressions of ethical? What are the key meta-ethical issues inherent in these extant notions (i.e., responsibility, fairness, justice, compassion, and rights)? Although we have formal legal definitions for many of these key ethical concepts, what are the tacit definitions held by educational leaders? The described notions of ethical held by these leaders remain unchallenged by traditional ethical analysis. In other words, now that the notions of ethical have been described, how do these understandings relate to the various historical and conceptual schools of moral philosophy? The notion of common life principles, for example, raises many theoretical questions. One might ask whether these common life principles are related to leaders' contextual roles or to universal principles? To what extent are "core" life principles unchanging and absolute, transient and relative, particular or universal? To what extent do educational

leaders reflect deontological, utilitarian, or natural law orientations in their understandings of the notion of ethical?

With respect to the leaders' definitions of the concept of ethical, it is apparent that educational leaders are relatively untouched by the sophisticated and technical nomenclature of moral philosophy. This is consistent with Bird and Waters' (1987) findings that leaders are not likely be systematic or traditional in their use of ethical language. Perhaps philosophers and practitioners would benefit from learning each other's language as a step toward growing in their appreciation of and access to each other's domain. I would suggest that the descriptions contained in this study provide a starting place for such undertakings. General ethical theorists may have a greater contribution to make when they have gained an understanding of the leaders' everyday world.

Implications for Research

The implications for research derived from this study point to a number of alternate directions that future research might take. As indicated at the outset of this article, the usual research approaches in applied ethics have been oriented toward the observation of behaviors or the analysis of participant responses to preconceived situations or ethical challenges. This article reports an alternative research approach that allowed educational leaders to voice their own understandings with respect to the notion of ethical. The point of this approach has been to determine some of the a priori conceptions of educational leaders with respect to the meaning of ethical. It has been said that,

> Administrators must develop skill in thinking about ethical problems, toward the end of creating a working professional ethic of their own. Without cultivating this ability to theorize and generalize from experiences, no public administrator can transcend the boundaries of particular events to comprehend and assess them. Without the illumination born of the marriage of abstract thought and practical experience, it is impossible to see where we are going. (Cooper, 1990, pp. 2-3)

I would suggest that applied ethics research in educational administration needs to be pursued on at least five broadly conceived fronts: descriptions of the ethical content of educational leader decision making (their problems, issues and conflicts); analyses of the normative ethical content of leader decision making (nature and relationship of formal ethical theory to leaders' commonsense ethical decision making); descriptions and analyses of the decisional processes (leaders' internal cognitions and affections in response to ethical challenges in education); descriptions and analyses of the contextual influences impinging on leader decision making (both exogenous and endogenous factors); and descriptions and analyses of the fundamental grounds used by educational leaders to resolve complex ethical conflict. This study represents only one dimension of the first approach (describing the extant ethical content of educational leaders, as related to their notions of ethical). Future research that thoughtfully considers the ethics of educational leaders, through any of the above approaches, will make a significant contribution to theory and practice in educational administration.

Implications for Practice

The implications of this study for practice may be cited in both general and specific terms. Every attempt must be made to demythologize ethics for the population. A number of the educational leaders involved in this study expressed a sense of intimidation with respect to ethics or, on the other hand, an apathetic attitude toward the subject. In general, the educational leaders in this study commonly expressed either an unrealistic set of ethical expectations (i.e., absolute ethical purity) or an extremely cynical view of ethics that held ethical integrity as an unattainable and unsustainable state. Practically, these polar perspectives might best be displaced by thoughtful and user-friendly approaches to the ethical domain. This might begin with the affirmation that within this particular professional group there was evidence of considerable ethical acumen. Rather than dealing with ethics exclusively from a professional discipline and policing orientation, the LEADS organization might instead develop a curriculum of ethics education for new members, including a more formalized system of ethical mentorship and networking. With respect to the professional disciplining of LEADS members as required by provincial statute, one might question the efficacy and practicality of an organizational code of ethics wherein members mean different things when they read words therein such as *ethical*. An implication of the finding that there is variation in the ways members think about ethics and wrestle with issues might entail incorporating this diversity into their collective ethics code.

I have observed that although no definitions for the notion ethical seemed to exclude positive ethical orientations of others, the wide range of understandings contributed to a general ethical malaise and seemed to encumber members from engaging in more ideal levels of moral conversation. Bird and Waters (1989), in their study of managers, consider the phenomenon of moral muteness among administrators. The implications of collective and individual ethical muteness, from a lack of definitional consensus, may indeed retard ethical development and result in fewer ethical initiatives by members with other members. Further, if the language of ethics is understood to be subject to great variance in interpretation, explicit consideration of the ethical domain in educational decisions may be less common. The advantages gained by collaborative deliberations on difficult ethical issues may be either diminished or enriched by this lack of definitional consensus. Those with influence in this particular professional organization may choose to celebrate, rather than mourn, this diversity of definition in order to encourage ethical dialogue and overcome ethical muteness. Individual educational leaders and their professional organizations, given such varied definitions, may also be predisposed to either operating with sententious or taken-for-granted definitions of ethical. There are, of course, many professional and organizational dysfunctions that attend to such tendencies. I would suggest that these findings point to the importance of educational leaders asking interpretative questions rather than assuming definitional consensus or embracing unfounded ethical positions.

In conclusion, perhaps the greatest practical implication of this descriptive research is its contribution to the particular educational leaders involved in the study. The findings provide a basis of information by which leaders may develop better understandings of, and learn from, the ways that their peers

conceive of the notion of ethical. The article provides definitions of ethical beyond descriptions of understandings (*begreifen*) that are imputed to leaders by deriving descriptions from leaders (*verstehen*) by listening to their voices. The purpose of this article is to offer readers an insight into the one aspect of the meaning of everyday ethical decision making for senior educational leaders through an alternate research methodology.

References

Ashbaugh, C., & Kasten, K. (1991). *Educational leadership: Case studies for reflective practice*. New York: Longman.

Bird, F., & Waters, J. (1987). The nature of managerial moral standards. *Journal of Business Ethics, 6*, 1-13.

Bird, F., & Waters, J. (1989). The moral muteness of managers. *California Management Review, 8*, 73-88.

Cooper, T. (1990). *The responsible administrator: An approach to ethics for the administrative role*. San Francisco, CA: Jossey-Bass.

Enns, F. (1981). Some ethical-moral concerns in administration. *Canadian Administrator, 20*(8), 1-8.

Foster, W. (1986). *Paradigms and promises: New approaches to educational administration*. New York: Prometheus.

Gronn, P. (1987). Notes on leader watching. In R.J.S. Macpherson (Ed.), *Ways and meanings of research in educational administration* (pp. 99-114). Armidale: University of New England.

Hodgkinson, C. (1991). *Educational leadership: The moral art*. New York: State University of New York Press.

Jackall, R. (1988). *Moral mazes: The world of corporate managers*. New York: Oxford University Press.

Ladd, J. (1957). *The structure of a moral code: A philosophic analysis of ethical discourse applied to the ethics of the Navaho Indians*. Cambridge, MA: Harvard University Press.

Leithwood, K., & Musella, D. (Eds.). (1991). *Understanding school system administration: Studies of the contemporary chief educational officer*. London, UK: Falmer.

Macpherson, R.J.S. (1985). Values, ethics and the Journal of Educational Administration. In F. Rizi (Ed.), *Working papers in ethics and educational administration 1985* (pp. 15-34). Australia: School of Education, Deakin University.

Maxy, S. (1991). *Educational leadership: A critical pragmatic perspective*. New York: Bergin and Garvey.

McCoy, C.S. (1985). *Management of values: The ethical difference in corporate policy and performance*. Boston, MA: Pitman.

Sergiovanni, T. (1990). *Value-added keadership: How to get extraordinary performance in schools*. Toronto, ON: Harcourt Brace Jovanocich.

Sergiovanni, T. (1992). *Moral leadership: Getting to the heart of school improvement*. San Francisco, CA: Jossey-Bass.

Smyth, J. (Ed.). (1989). *Critical perspectives on educational leadership*. New York: Falmer.

Toffler, B.L. (1986). *Tough choices: Managers talk ethics*. New York: Wiley.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 35-56*

R.J. Carney

and

H.W. Hodysh
University of Alberta

History of Education and the Rite of Passage to Teaching: The Alberta Experience 1893-1945

The history of education was among the first subjects to be studied by those preparing to teach in the Province of Alberta. As such the subject was an early manifestation of attempts to relate theoretical principles of teaching to educational practice. Its significance in the normal schools and the University of Alberta from 1893 to 1945 is explored with reference to the subject's formative role in the transitional stage of Arnold van Gennep's concept of The Rites of Passage. *Explored is the central hypothesis that teacher educators saw the history of education as a core element in providing a unique body of professional values and standards to those wanting to teach in the province's publicly supported schools.*

Autrefois, l'histoire de l'éducation était un des premiers sujets qu'étudiaient ceux qui se préparaient à enseigner dans la province de l'Alberta. Tel qu'il était conçu, ce sujet peut être aperçu comme étant une manifestation hâtive qui essayait de relier les principes théoriques de l'enseignement à la pratique éducative. Son importance dans les écoles normales et à l'Université de l'Alberta entre 1893 et 1945 est explorée en faisant référence au rôle formatif de ce sujet dans les stages transitionnels du concept énoncé dans Les rites de passage *de Arnold van Gennep. On explore l'hypothèse centrale que ceux qui formaient les enseignant(e)s perçevaient l'histoire de l'éducation comme étant un élément central pour munir les futurs enseignant(e)s qui voulaient enseigner dans les écoles publiques de la province, un bagage de connaissances normalisées et de valeurs professionnelles.*

When Carr (1987) suggests that institutional values shape our understanding of the past, he is indicating their effect not only on the process of historical inquiry, but perhaps more importantly on how values emanating from that process have influenced our understanding of events. Most historians of education, until recently at least, would agree with Carr's observation, noting that in the process of teacher education those values emerging from such study have had a significant effect on the theory and practice of teaching. As is shown, these remarks have particular relevance for the Province of Alberta from the late 19th to the mid-20th century where the normal schools and the University of Alberta shared value judgments about the role of educational

Robert Carney is a professor in the Department of Educational Foundations. His research interests include the history of schooling in the North, intercultural education, and the history of Canadian education. He is currently writing a monograph on residential schools in the Diocese of Mackenzie-Fort Smith.

Henry Hodysh is a professor in the Department of Educational Foundations. He has published widely in the field of educational history where questions of historical theory and the uses of the educational past remain a continuing focus of his research.

history in teacher training. To use van Gennep's (1960) analysis, history of education[1] became an integral element of a rite of passage to teaching.

Published as *Les rites de passage* in 1908, van Gennep's examination of the values and rituals associated with such major personal changes as puberty and marriage in tribal societies subsequently influenced research on social status and professional socialization (Chapple & Coon, 1942). This was particularly evident in studies of teacher education and training[2] after the work was translated into English in 1960. An early use of van Gennep's concept can be found in Fuch's (1969) *Teachers Talk*, where the beginning teacher's first formal evaluation is identified as the most significant event in the rite of passage to teaching. Subsequent teacher training research, which used the rite of passage as a model, has frequently assigned similar significance to evaluations of teaching during the training period (see, e.g., Eisenhart, Behm, & Romagnano, 1991; Gehrke, 1991; Hale & Staratt, 1989; Sylwester, 1987; Tinto, 1988; White, 1989). Even those like Lacey (1987) who regret the lack of research on later stages of teacher socialization recognize the importance of the initial teacher formation period:

> The middle ... years of professional life have been almost totally neglected. The reasons for this are fairly clear. The training period is the stage where the neophyte professional is being consciously shaped by others, the trainers. There is, therefore, an interest in the outcome of this process because in theory it can be altered to produce "better" or different results. The researcher is also drawn to study this period because it is a stage of massive personal reorientation and change, a *rite of passage*. (p. 637)

Most research on this question emphasizes methods and evaluations of classroom practice. An example can be found in White's (1989) "Student Teaching as a Rite of Passage" where she contends that the body of cultural knowledge gained by neophyte teachers through mastery of classroom routines is an essential step in achieving teaching competence. She makes it clear, however, that the acquisition of "several basic tenets about teaching," should not be grounded on feelings or impulse, or, as some critics of teachers' knowledge would have it, on a "mixture of idiosyncratic experience and personal synthesis." What needs to be pursued, in her opinion, is "the study of a cultural body of knowledge [about teaching] that is transmitted and acquired from generation to generation" (p. 192). Both White and Eisenhart et al. (1991) lament the tension that exists between teaching theories and teaching practices. In the view of Eisenhart et al. this tension has led to a tendency among student teachers "to disregard the university as a source of information about teaching" (p. 66). One might conclude from such studies that theories of teaching in history of education courses would have no place in the rites of passage to teaching advocated by White (1989) and others. On the other hand, one could infer the opposite possibility, given White's call on the teaching profession "to ensure that its version of reality and in particular its fundamental tenets will be accepted and transmitted" (p. 178). Her affirmation evokes what appears to be a longstanding belief among those responsible for teacher education in Alberta; namely, that the study of educational history makes a significant contribution to a unique body of professional values, knowledge, and practice.

The central hypothesis of this study is that the purpose of the history of education, as a core element in the rite of passage to teaching in Alberta from the 1890s to the 1950s, was to provide beginning teachers with a set of professional values and standards. In the discussion that follows it is shown that the values and norms transmitted in history of education courses were based, as Harvey (1986) would argue, on a universe that was viewed as being objective and unchangeable. Those values were reflected in the purpose and content of teacher training and education. What, then, was the nature of educational history as a field of study at the normal schools and the University of Alberta? And how as a rite of passage did its development reflect underlying value judgments about standards or principles of worth in the education of teachers? (compare Kaplan, 1964, p. 370).

Background

The history of education as a key subject in teacher training in English-speaking Canada is evident as early as the 19th century in the normal schools of Ontario and elsewhere, and in the 20th century in university-based teacher education. This was particularly so in Alberta, where most teacher educators have supported the view that the study of educational history gives a theoretical, value-related basis for an understanding of the nature and methods of education in the schools, providing, as it were, "a cultural body of knowledge" about teaching (White, 1989, p. 192). According to van Gennep (1960), a rite of passage—such as the movement from adolescence to adulthood—involves three distinct stages: separation, when one leaves a specific social state for another; transition, when one is presented with the values of the new state; and incorporation, when one is deemed to have adopted a new social status.

Most references to teacher education as a rite of passage emphasize the credentializing outcomes of teacher evaluations that signal an end to the process of incorporation. The initial stage of this passage in many instances involves separating neophyte teachers from their families, friends, and other students. Other discontinuities include new social networks, unfamiliar duties, and different language codes. During the transitional stage of the rites process, institutional authorities attempt to overcome the initial difficulties and sense of isolation of novice teachers by reorganizing "their behaviour, appearance, speech and ways of thinking" (Eisenhart et al., 1991, p. 54). The transitional period, then, is one in which the novice is instilled, by means of precept and practice, with a set of largely conventional establishment-oriented educational values.

The history of education has played a significant role in training teachers, principally during what rites of passage theorists call the transitional stage. There were three purposes behind the inclusion of the subject in the curricula of the early normal schools. In the first instance, the history of education provides a separate canon of historical knowledge on the theory and practice of teaching that helped justify the existence of distinctive forms of teacher education. Second, teacher educators expected that teachers-in-training would incorporate relevant ideas from the works of great educators in their studies and especially in matters relating to classroom practice. Third, it was thought the knowledge and use of the historical canon would lead to the transformation of teacher candidates into advocates of public schooling and the existing

social order. The extent to which these purposes have been achieved at any particular time has occasioned much debate. The belief among teacher educators that the history of education and related foundational studies, like philosophy and sociology of education, are invaluable teacher training program components persists, even though students have found them to be primarily of theoretical interest (Koos & Woody, 1919; Lacey, 1977; Miklos, Conklin, & Greene, 1987; Pettifor, 1948).

Although the seeds of this development can be traced to the classical period in western European education, particular reference might be made to the *licentia docendi* of the medieval university, which gave recipients the right to teach anywhere in Europe (Rashdall, 1936, pp. 278-284). Based on the trivium and quadrivium of the seven liberal arts, this license for teaching indicated the control of university instruction by academics as a professional group (Kimball, 1986; Le Goff, 1980). The license, however, reflected more than a knowledge of subject matter. It implied a moral suitability based on an evaluation of the scholar's life and reputation (Ferruolo, 1985). Although academic jurisdiction remained a contentious point of debate between secular and ecclesiastical interests, the *licentia docendi* was firmly in place by 1231 (Le Goff, 1980).

In addition to being continually subject to the scrutiny of the university community about what should be taught, advocates of teacher training were faced with the task of determining how good teaching practices might be learned. Attempts to address this, for instance, were manifested in the work of the Society of Jesus founded in 1534. Given to an interest in secondary education, the Jesuits have been credited with introducing the practice of teacher training, which in their view involved a uniform pattern of instruction based on study and observation (Eby, 1952). The Jesuit program, including the study of languages, literature, philosophy, theology, and mathematics as well as teaching methods, stressed self-discipline and religious and moral training (Schwickerath, 1903), with a sound knowledge of "general culture" serving as the basis for specialized training. For the Jesuits the study of history by way of original sources was viewed as critical for "argument, illustration, and parallel" (Farrell, 1938, p. 249). As in all Jesuit pedagogy, emphasis was placed on understanding, and then on developing mastery (Donohue, 1963).

Other examples might be used to illustrate an attempt to instill a professional awareness of teaching in the early history of education, but of special interest are the contributions of the Christian Brothers established in 1684 by Jean Baptiste de la Salle. Focusing on the primary education of working class children, de la Salle established what has been identified as the first normal school (Eby, 1952). Neophyte teachers were prepared in the simultaneous method for instructing large classes of children with emphasis placed on moral and literary studies as a preparation for life (Battersby, 1952).

As Kimball (1992) suggests, the idea of teaching as a profession originated at the time of Cicero. It involved a public declaration (*professio*) of a vocational commitment. By the period of the high Middle Ages, the idea denoted one having made a "profession of religion" by taking "the vows of some religious order" (p. 19). As many clergymen, like the Dominicans and later the Jesuits and Christian Brothers, were also teachers, they became responsible for directing other church members in matters relating to formal education. The value

orientations of this instructional ideology were rooted in theology (the queen of the sciences), which was the special preserve of the clergy. Similar educational orientations applied to Protestant communities, such as those prevailing in Puritan areas of colonial America. The concept of profession was subsequently expanded to include "the fields with which theology had been associated since the thirteenth centuries, law, medicine and, learning" (Kimball, 1992, p. 302). With the advent of state-sponsored national systems of education in the late 18th and early 19th centuries, however, the clergy's influence on teachers became less pronounced. Not enough professed teachers were available to control the burgeoning number of common or public schools. Moreover, as the preparation of teachers for these institutions became the responsibility of the state, the close relationship between the clergy and the task of teaching was no longer possible. As a result, new forms of teacher socialization were needed that would include a set of common values. The expectation in the nonsectarian normal schools established for this purpose was that these values would be consistent with the institutional goals of public schooling and, at the same time, would give teaching the dignity and the independent status attained by other professions.

These early developments in history and teacher education are reflected in attempts to establish permanent institutions for teacher training in mid-19th-century Ontario and the Maritimes. In 1858, for example, Angus Dallas, an advocate of private and denominational education, deemed the work of the Toronto normal school, which had been established in 1847, to be an "expensive fraud" (Phillips, 1957, p. 571; compare with Houston & Prentice, 1988, pp. 167-169). Nor was the idea that secondary school teachers needed special training a popular one. Universities tended to regard "professional training as an insult to their arts and science graduates and as an implied criticism of their own teaching methods" (Stamp, 1982, p. 44). For example, C.F. Lavell of Queen's University claimed that anyone who argued that high school teachers needed training was "a scarcely endurable heretic" (p. 44). These critics invariably argued that appropriate teaching practices could be gained by having prospective teachers emulate the sound pedagogical practices found in schools and universities. In their opinion, candidates for teaching could determine which types of instruction would have the desired effect simply by witnessing existing teaching methods and by recalling examples of good teaching practices from their own schooling.

The debate over how best to achieve curriculum competence and appropriate teaching strategies remains contentious. What was clear in the minds of most teacher educators well into this century, however, was that history of education would provide the principles and standards for resolving a number of fundamental professional questions, such as what knowledge, skills, and attitudes were of most worth and how best they might be imparted. It is not surprising, therefore, that the subject figured prominently in the initial course of studies at the Toronto normal school. From that time onward, it became one of the key professional courses in the normal school curricula, even though, as Alexander Forrester (principal of the Nova Scotia normal school from 1855 to 1869) put it, not enough instructional time was available for this important area (Phillips, 1957, p. 576).

Yet even when educational history was taught, its influence was not always apparent. For instance, L.M. Montgomery (1985), who attended lectures on great educators at the Charlottetown Normal School in 1893, expressed definite ideas about "good" and "bad" teachers in her best-selling *Anne of Green Gables* series (p. 96).[3] For Anne Shirley, the main character in the novels, teachers were seen in terms of polar opposites: "The Good Teacher has no faults; the Bad Teacher no redeeming qualities" (Gates, 1989, p. 116). Such archetypes go back to the childhood of the author and apparently had little to do with her time at normal school. Gates states that "in the rites of passage ... of becoming a teacher," Anne lost this childhood view of teaching for "a more adult understanding of human beings," largely as a result of her first teaching experiences (p. 169). Gates' statement is in keeping with other rite of passage commentators who view evaluations of teaching practice as a critical element in the process (see, e.g., Eisenhart et al.; Gehrke, 1991; Hale & Staratt, 1989; Sylwester, 1987; Tinto, 1988; White, 1989). For many historians of education, this emphasis on evaluation overlooks what "better teachers" might have done in such circumstances. After punishing one of her pupils, Anne reflects as follows on this perennial issue of theory versus practice:

> "I never expected to win him by whipping him, though," said Anne, a little mournfully, feeling that her ideas had played her false somewhere. "It doesn't seem right, I'm sure my theory of kindness can't be wrong." (Montgomery, 1981, p. 102)

Theories of good teaching abounded in the normal schools of the 19th century and involved lofty ideals of what student teachers could accomplish. The development of teacher training in Upper Canada (Ontario) coincided with Egerton Ryerson's tenure as Superintendent of Schools (1844-1876). On taking office Ryerson was given a year's leave to examine state-supported educational systems in Western Europe and the Eastern United States. Nothing occurred during his travels to change his political perspective and the implications it had on matters relating to schooling and teacher training, evidenced in his remark (Thomas, 1969) that "my leading idea had been ... to render the Educational System, in its various ramifications and application, the indirect but powerful instrument of British Constitutional Government" (p. 100) The course of study at the Toronto normal school, which included some references to the history of education in the late 1850s,[4] and which by 1882 involved lectures on "great Educational Reformers and their Methods" (Ontario, 1882, p. 76), was a model for professional courses and teacher training programs subsequently developed in the Canadian West.

Educational History and Normal Schooling in Alberta

D.J. Goggin, whose reputation as an outstanding educator included service as head teacher of a model school in Ontario, was appointed principal of the Protestant normal school in Winnipeg in 1886. He remained in this position until 1893, when he moved to Regina to become the first director of normal schools and teacher institutes in the Northwest Territories. Shortly after assuming this position, he became Superintendent of Education for the Territories, an appointment that strengthened his control of teacher education arrangements. The Regina normal school's syllabus adhered closely to the

teacher training program that Goggin had developed in Manitoba. In 1896, for example, normal school students in the Territories who wished to obtain a second- or first-class teaching certificate had to complete a course in the history of education (North-West Territories, 1897, pp. 18-19). The prescribed textbooks for the second class requirement in history were Browning's (1881) *An Introduction to the History of Educational Theories* and Quick's (1895) *Essays on Educational Reformers*, whereas those for the first-class were Spencer's (1911) *Essays on Education and Kindred Subjects*, *The Report of the Committee of Ten on Secondary School Studies with the Reports of the Conferences Arranged by the Committee* (1894), and Rosenkranz's (1890) *The Philosophy of Education*.

The inclusion of the text by Herbert Spencer, a noted English philosopher and sociologist, marked the introduction of works of a philosophical nature into the normal school's history of education syllabus. As a result the field of study became a hybrid of historical and philosophical orientations that principally encompassed references to the history of educational ideas. It mattered little, therefore, whether courses in this area were titled the history or the philosophy of education, or whether they were offered separately or in conjunction. Historically oriented textbooks, such as those by Browning and Quick, tended to describe the thought of past educators, whereas texts of a philosophical cast, such as Spencer's, took a prescriptive, albeit general, approach to matters related to teaching. References such as the latter, which at the time were considered to be philosophical in nature, were also much more given to applying a particular theoretical perspective to contemporary educational issues.

Although the prescribed texts for the 1896 history of education classes provided suggestions for educational change, none linked the thought of a particular educator to a system of classroom practice deemed worthy of emulation in most respects. One has to turn to earlier texts in the history of education, such as Forrester's (1867) *The Teacher's Text Book*, to find individuals singled out as having provided an exemplary teaching methodology. Like Goggin some 30 years later, Forrester was a superintendent of education and a normal school principal, but differed from Goggin by advocating a system of teaching developed by someone who had extensive classroom experience, and whose success as a practitioner meant much more than theories about what constitutes good teaching. In designing a program for the first normal school in Nova Scotia, Forrester examined five approaches to teaching: rote, mechanical, explanatory, objective, and training. After briefly outlining the principles and applications of each approach and placing it in historical context, Forrester (1867, p. 308) concluded that "The Training System" developed by David Stow, an early 19th-century Scottish educator, to be the one to follow. Although predicated on the scriptural rule of rearing a child in the way he should go, Stow's system had none of the harshness usually associated with this dictum. It was, in fact, one in which neither rote learning nor physical punishment had a place. And although the system's intricacies and merits cannot be discussed here, it should be noted that Forrester's choice of a model educator and his efforts to show the classroom implications emanating from his choice differed substantially from the objectives and content of the history of education offerings at the Regina normal school and its successors.

Until the 1930s teacher candidates at these institutions were provided with a summary of the writings of eminent educators from antiquity to current times. This was achieved in part by including prescribed texts of a philosophical nature in history of education courses. As noted earlier, Rosenkranz's *The Philosophy of Education* became one of the required readings in 1896 for the first-class course in the discipline at the Regina normal school (North-West Territories, 1897, p. 20). Like Spencer's text, it was a work that set out a theoretical perspective for students. The expectation was that in becoming familiar with what were deemed to be appropriate educational ideas from the past, prospective teachers would form them into a conceptual matrix for emulation and practice. Access to the work of contemporary philosophers such as Spencer and Rosenkranz was seen as a way of facilitating this exercise. Such were the underlying assumptions that Goggin set for the history of education in the 1890s. Following his approach and the philosophical ideas of Rosenkranz and Spencer, the questions of the 1898 Regina normal school examination on the history of education were wide-ranging and eclectic in nature, emphasizing theory and educational ideals in classical, medieval, and modern-day contexts (North-West Territories, 1898). By way of comparison, the 1903 examination called for a more interpretive understanding of the educational past, and went beyond the descriptive-type expectations of earlier examinations (North-West Territories, 1903).

According to the 1903 course of studies at the Regina normal school, history of education also encompassed topics in psychology and the philosophy of education. The schedule and content of the examinations for that year reveal, however, that the history of education was losing its omnibus character and preeminent status as the key science of education subject. In addition to the history of education test cited above, there were separate examinations in psychology and philosophy. Those in the latter area were two and a half hours in length compared with the hour and a half given to history. A review of the examination questions in the three subjects also suggests that the school's determination to relate educational theory to classroom practice was achieved more successfully in the newly separated disciplines of psychology and philosophy (North-West Territories, 1903).

A change in history of education textbooks occurred following the establishment of the Province of Alberta's first normal school in Calgary in 1906. Quick's and Browning's texts, which given their British perspective were linked to the cultural and political predispositions of educators like Ryerson and Goggin, were replaced by an American text, Painter's (1901) *A History of Education*, which extolled the idea of public schooling. This change diminished the possibility of relating schooling to existing Anglo-conformist views of what the ethos of the province' schools should be, thereby making the discipline more remote from the day-to-day activities of teachers than had previously been the case.

Painter's book manifested many of the characteristics that Bailyn (1960) ascribes to other American history of education texts of the period, including Davidson's (1900) *A History of Education*, which was listed as a reference at the Calgary normal school in 1907 (Alberta Department of Education, 1908, p. 104). According to Bailyn (1960), Davidson's book was designed to give "the

neophyte [teacher] an everlasting faith in his profession" (p. 8). As with other historians of education in the United States at the time, Davidson (1900) set out to provide an overview of education, encompassing the rise of human intelligence through various epochs and its evolution to its highest place in the profession of education. The history of education thus conceived was an initiation to "the noblest of professions" (Bailyn, 1960, p. 7). The theme of the discipline's role as a rite of passage abounded in other works of the period, such as Monroe's *A Text-book in the History of Education* (1905 reprinted in 1930). Bailyn chose the following passage from Monroe's work to describe what "a whole generation of passionate crusaders for professionalism in education" had in mind for the student of the history of education:

> a conception of the meaning, process and purpose of education that will lift him above the narrow prejudices, the restricted outlook, the foibles and the petty trials of the average school room, and afford him the fundamentals of an everlasting faith as broad as human nature and as deep as the life of the race. (p. x)

The initiatory role assigned to the history of education was not new. Transition into teaching had been a function of the history of education in state-supported normal schools in North America almost from their inception in the mid-1800s. And as the following excerpt from the Ontario Ministry of Education *Report* (1891) illustrates, the subject gained increasing significance over the decades:

> The knowledge of what has been done by the great educators of the past, their educational ideals, their modes of procedure, their strengths and sacrifices and triumphs, animate the teacher with loftier ideals and with the spirit of effort; the critical study of the underlying principles of various national systems, their truths and their errors, helps at once to enlarge and make clearer our ideas of the science of education. (pp. 426-427)

The report also affirmed that the history of education, along with other principles of education, including those derived from psychology, would result in a "change of spirit and purpose" in future teachers, not unlike that experienced by someone who had undergone a "*conversion*" (p. 427).

What differed about the orientation of educational history in the early 20th century was that those responsible for the field of study either lessened or abandoned attempts to relate its theoretical orientations to classroom practice. As far as the normal schools in Alberta were concerned, there is evidence that as the theoretical orientations of the subject gained ascendancy, linkages between educational theory and specific coping and instructional strategies were left to other disciplines such as psychology. These orientations were strengthened further, as was the case in 1908 at the Calgary Normal School, when the history of education and the philosophy of education were amalgamated into a common course for first-class students only (Alberta Department of Education, 1908). Involving a series of reflections on selected educational ideas, the common course continued at Calgary and was adopted by the Camrose Normal School when it opened in 1913. Apparently there was uneasiness about the absence of a theoretical course in either or both of the disciplines for second-class candidates, but this was rectified in 1909 when a course in the history of education was prescribed to ensure that students at this

level were properly initiated into the profession (Alberta Department of Education, 1909).

Although there is nothing to indicate how these courses were viewed by the students, it would have been unlikely for them to publicize any misgivings they might have had about the subject, especially when the principals were usually responsible for teaching courses in this area. Normal schools were under the direct control and supervision of the provincial government's department of education and as such did not have the independence in curriculum and other matters enjoyed by the University of Alberta. What should be noted, however, is that educational authorities in Canada and the United States, including department officials and principals of the normal schools at Calgary and Camrose, were generally of one mind about educational history being part of the professional training of teachers. The National Educational Association in the United States recommended this in 1907 (Cremin, 1955), and as has been shown above, teacher educators in Alberta took special pains to see that the subject remained part of the core curriculum. A study in the State of Washington in 1917 found that 315 of the 451 teachers surveyed had taken educational history as part of their professional training. Only 38 of the 315, however, said it was "their 'most helpful' course, ranking it one step above [the] elementary school curriculum, the subject stated by the teachers to be the least value" (Koos & Woody, 1919, pp. 246-257). It is not known if the Alberta normal school principals were aware of the Washington survey, but one might conclude that they would have believed that their educational history courses would not be similarly assessed.

In 1919 the normal school course was extended from four to eight months. To meet the expected shortage of teachers that would result from the longer program, an emergency short-term course of 12 weeks leading to a third-class certificate was established in Edmonton. No instruction in the history of education was given in the short program, an exclusion that would have concerned those who believed in the subject's value as a rite of passage toward teaching. The short course was offered for one term only. From then on, the Edmonton Normal School followed the same basic course of studies as the two other normal schools. Educational history remained a requirement for second- and first-class candidates, but the first-class course no longer involved studies in the philosophy of education. As a result, texts like Horne's *Philosophy of Education* (1905) and Tompkins' *Philosophy of Teaching* (1894) were dropped from the prescribed and recommended lists of readings, for plainer fare in the form of such works as Monroe's *Text-book in the History of Education* (1905) which provided the content for the two courses in the history of education (Alberta Department of Education, 1909, p. 102). According to the Department of Education *Normal School Announcement* of 1918-1919 (Mann, 1961), this involved a "brief survey from ancient to modern times" as the first-class level and "lectures on the most important movements and problems" at the second (pp. 72-73).

That educational history was usually taught by the principals of the normal schools, who also played a major role in issuing teaching certificates, would not only have enhanced its stature, but would likely have led, in keeping with most ceremonies of transition, to a suspension of judgment on the part of the

students about the future value of the experience. Once the course began to be taught by regular members of staff in the 1920s, such as C. Sansom and A.G. Torrie at Calgary, it became like any other offering of the program and tended to be judged on its own merits. Sansom and Torrie went on to become principals of the normal schools at Edmonton and Camrose respectively. On taking these positions, however, they gave up their involvement with the history of education to devote themselves to subject areas such as mathematics and psychology, which were apparently of greater interest to them.

The reports of the provincial normal schools from the 1920s to the mid-1930s (Alberta Department of Education, 1921-1935, *Annual Reports*) made little mention of the history of education. According to Mann (1961), it continued to be taught to 1937 for up to three periods a week, about 10% of the instructional time, but was available to first-class students only. In this regard it is worth noting that Black (1936), who studied teacher education in Western Canada in the 1930s, found it "difficult to understand why the Alberta authorities have made this distinction; the second-class students are probably just as much in need of a historical background as are the first-class students" (p. 112). By this time Monroe's text had been replaced by Graves' *A Student's History of Education* (1925).[5] This book was only slightly more suitable given the course's new title, History of Modern Education, and its description in the 1936 summer session calendar:

> Educational philosophers of the present day and their work. Changes in the conceptions of the process of learning and the influence of these in both subject matter and teaching practice. The "New Education." Promising experiments. Present day trends in education. (Alberta Department of Education, 1936, p. 29)

The course's restructuring probably occurred largely as a result of a department-sponsored report in 1935 by two normal school instructors, Donalda Dickie and Olive Fisher, and an inspector of schools, Hay, who recommended that an "enterprise program" be introduced in grades 1 to 6. Enterprise units were built "on the principle that education is a social experience in the course of which pupils plan, initiate, and carry out cooperative projects" (Alberta Department of Education, 1935, p. 18). At a conference of normal school instructors in 1935, chaired by H.C. Newland, provincial supervisor of schools, agreement was reached on a number of matters relating to the implementation of the enterprise. Of significance here was a recommendation that called upon the normal schools to give "more attention ... to the philosophy of education" so that teachers would better understand "the meaning and purpose of the changes in the curriculum" (Alberta Department of Education, 1935, pp. 20-21).

Newland, who had been one of Goggin's students at the Regina normal school, expanded further on what was needed in a report on "The Training and Certification of Teachers for Alberta Schools" that he sent to William Aberhart, premier of the province and minister of education, in June 1936. Newland proposed a new normal school curriculum consisting of 20 periods a week, two of which would be assigned to a course on The Principles of Education. The new course would provide a "very brief outline" of educational history, but most of its attention would be devoted to such contemporary topics as these:

> The role of education in a changing society; education and democracy; the liberal, cultural, technical and vocational objectives of education; curriculum-making; the child-centred and subject-centred school; the functions of the elementary school, the intermediate school, and the high school; educational problems in Alberta. (Newland, 1936, p. 34)

There is no record of Premier's Aberhart's reaction to what was proposed. Suffice it to say that energetic steps were taken to introduce the enterprise program and allied courses in summer school classes and during regular normal school sessions. According to departmental officials like Newland, new approaches were needed, such as that taken by G.K. Haverstock, principal of the Camrose Normal School in 1936, when he did away with educational history entirely and replaced it with a course in Elementary School Organization and Management (Newland to G.S. Lord, 7 December 1936, *Teachers College*).[6] Similar initiatives led to the history of education being removed from the curriculum of normal schools a year later in favor of a course on The Principles and Science of Education (Mann, 1961, p. 25). This offering, which was designed to deal with the new activity or enterprise program, was further reinforced by similarly oriented courses during World War II, such as Community Economics and Sociology and Community Problems (p. 251).

The enterprise program had its general origins in the work of John Dewey, William Kilpatrick, George Counts, and others who formed a loose coalition that was known as the new or progressive education movement. Its principal ideologue in Alberta was H.C. Newland who adopted many "progressive" ideas and whose writing and speeches about their educational implications were invariably enthusiastic. As he put it in 1924, once teachers became aware "of the possibilities of education as the harbinger of a new and better social order," they would become "apostles of progress and not reaction" (Patterson, 1974, p. 293). At a summer school for teachers in the 1930s, Newland made the following remark on progressive education: "We are on our way. We don't quite know where we are going. But, a good thing is we can't go back" (H.T. Coutts, personal communication, 18 September 1992). There can be no doubt that Newland's proposals would not have gained the influence they did in the 1930s without the commitment and cooperation of Donalda Dickie, W.E. Hay and D. McDougall who showed remarkable ingenuity in translating progressive ideas into classroom practices. The theory and practice of the enterprise had a much greater effect on the thinking and actions of teachers in the province than any of the transitional efforts that had been associated with educational history. As far as this latter subject was concerned, it lost its place in the core curriculum and all but disappeared in other aspects of the normal school program.

Although educational history was excluded from the core curriculum, education officials were not disposed to do away with the discipline entirely. They saw it playing a new role as part of classroom-oriented courses where it took a critical, even iconoclastic stance concerning the nature and outcomes of public schooling in particular and socioeconomic conditions in general. Attempts to do this are apparent in such courses as the Social Foundations of Education which was first offered during the 1936 summer session and which was described as follows in the sessional calendar:

> This course has been designed to show the social purpose of education, and to help teachers understand the social philosophy on which our educational system is based. It will deal ... with the effect on education of the change from an agrarian to an industrial economy. It will also examine the role of education in a changing society, and the bearing of education on the future of democracy. (Alberta Department of Education, 1936, p. 30)

The instructor for the above course used Counts' (1932) *Dare the School Build a New Social Order?* as a principal text in the University of Alberta's summer session in 1936. The book's contemporary nature and radical politics was a major departure from previous history of education perspectives.[7] William Swift, a school inspector who became Deputy Minister of Education in 1946, was asked to give two other hastily designed courses in the 1936 summer session: Rural Sociology and Rural Education and its Problems. He recalls being ill prepared to teach either course, and describes the Department of Education's expectation that his courses should be related to the enterprise program to be introduced in the schools that fall as "muddled" thinking (W.H. Swift, personal interview, October 29, 1992).[8] Like many of his colleagues, Swift was nonetheless committed to introducing the enterprise, and by association at least, to many of the socioeconomic outcomes that progressive theorists believed would result from it and other educational innovations. What the above courses and related information reveal, however, is that the new history of education was no longer seen to be in league with the existing sociopolitical order. Although the discipline ostensibly retained its transitional role in preparing teachers, its function as a rite of passage to teaching changed from one of accepting and reinforcing what prevailed in schools to one that was given to criticizing and redirecting what went on in these institutions.

Those in the normal schools and elsewhere who still saw the history of education as having significant value in the training of teachers undoubtedly hoped that the University of Alberta, which assumed responsibility for all teacher education in the Province in 1945, would not follow the same route as the normal schools and abandon a subject that had been so carefully nurtured in teacher preparation programs for most of the previous century.

Educational History at the University of Alberta

The value of educational history as a rite of passage at the University of Alberta can be traced to a series of events in the first half of this century. As early as 1911, H.M. Tory, President of the University, called for "a philosophical course for Education in keeping with the Historical training for Law and the Science training for Medicine" (University of Alberta Senate, 1911). Implicit in this view is the belief that familiarity with the thought of great educators was an essential element in the transitional and incorporating process of teacher training. Tory was among the first to call for educational studies in the fledgling Bachelor of Arts Program for those interested in teaching in the secondary schools. A combined study of educational history and philosophy was introduced in the Philosophy Department in 1912.

The History of Education, a senior level course, was given three hours per week throughout the 1912-1913 academic year. In a two-to-one ratio of history to philosophy, the course included historical studies of ancient, medieval, modern, and contemporary education with philosophical studies addressing

"educational ideals and social aspects of education" (University of Alberta, 1912-1913, pp. 71-72). The first examination in the subject centered on questions of history and educational thought, including a comparison of Greek and Roman ideals of education, key ideas in the works of William James, and an account of the origins and development of the universities in Europe (University of Alberta, 1913-1914, p. 71). The course was taught by the Rev. Dr. S. Dyde, an experienced teacher educator, who later became principal of the Presbyterian Theological College at Queen's University in Kingston. In addition to lectures, the recommended text in the course was Monroe's *Text Book on the Theory of Education* (University of Alberta, 1914-1915, p. 117). By 1915, History of Education was retitled Education 51, History and Philosophy of Education (University of Alberta, 1915-1916, p. 143).

These early developments in educational history are given a sharper focus with the founding of the School of Education within the Faculty of Arts in 1928-1929. Under the aegis of the Joint Committee of the Provincial Department of Education and the University, whose membership included among others W.A. Kerr, J.M. MacEachran, G.F. McNally, and M.E. LaZerte, it was recommended that along with studies in psychology, science methods, and administration, Education 54, History and Philosophy of Education should be included in the "professional subjects of the teacher-training course" (Joint Committee, 1929). The studies, which were designed to provide "a systematic and cohesive preparation in scope and subject matter for the work of teaching in the schools of the province" (University of Alberta, 1929-1930, pp. 91-92) indicated functional objectives of, first, transmitting to the rising generation the world's accumulating store of knowledge and, second, of assuring in close cooperation on par with those already provided in the "sister professions" of agriculture, engineering, law, and medicine (p. 92).

By 1935 the School of Education had already established two courses identified with the history of education. The first, Education 54 taught by H.E. Smith and M.E. LaZerte, focused on a study of classics and philosophy of education, Education 101, History of Educational Administration, centering on an historical study of educational administration in various countries (University of Alberta, 1935-36, pp. 153-154), a subject that disappeared from the curriculum four years later. Soon to follow in 1936 was Education 104, History of Education taught by H.E. Smith. This course addressed "the most important events in the history of European education and their effect on present-day practice in Europe, the United States and Canada" and included topics from the early Greeks to pioneers of modern education, as well as the emergence of national school systems (University of Alberta, 1936-1937, p. 158).

In 1939 the School of Education realigned its degrees by assigning the direction of the MA in Education and the BEduc, both of which were advanced degrees, to the School of Graduate Studies (Myrehaug, 1972). These changes, along with influence from the teaching profession, led to the formation within the Faculty of Arts of the College of Education in 1940, organized "for the training of high school teachers and for research in the field of education" (University of Alberta, 1940-1941, p. 235). Included in the BEduc curriculum was Education 54. Later, when the MEd degree was instituted, among the

requirements were Education 54 and other courses, such as Education 104, History of Education.

Of special interest are the topics of research in educational history, four of which are given particular attention. In 1915 A.E. Ottewell, under the supervision of Dr. MacEachran, presented a thesis on *The University Extension Movement*, an historical survey of university extension in Britain, the Antipodes, the United States, and Canada. Ottewell's study included extensive reference to historical and current events bearing upon extension activities. Following Ottewell's investigation little was done by way of thesis research into educational history until 1938 when K.H. Thomson examined the development of Russian education during the time of Leo Tolstoy. *The Educational Philosophy of Tolstoy* centered on an historical analysis of Tolstoy's life with attention to his views of education. In the following year Thomson (1939) completed a BEduc thesis—at that time a graduate degree—on *The Educational Philosophy of Robert Owen*. As in his earlier study, Thomson focused his research on the individual's contribution to education with attention to educational thought.

These initial studies were followed in 1940 by Hewson's *The History of Commercial Education in Canada*, a wide-ranging thesis on commercial education by provinces. Reference is made to population growth, economic change, and programs of study throughout the Dominion based in part on primary source materials including government annual reports. It is difficult, of course, to determine the impact of four dissertations on research in the history of education, yet two observations may be appropriate: first, although the studies were prepared in the Faculty of Arts, they had a specific focus on events and individuals in education; and, second, the topics addressed what is frequently identified as the history of educational ideas and, especially in Thomson's work, the history of educational thought. Whether or not these perspectives followed what already had been established in European educational history (compare Hodysh, 1980), they provide a view of educational history that included an interest not only in the study of schooling and education in the wider sense, but the philosophical and educational thought of leading historical figures.

The early development in educational history at the University reflects underlying commitments to both the professional training of teachers beyond the normal school and the transmission of knowledge about the western educational tradition, defined in part through European and American experiences. These commitments indicated a series of value judgments about "the standards or principles of worth" (Kaplan, 1964, p. 370, and compare with Gauss, 1990, p. 219; Kuhn, 1977, pp. 335-336) in the history of education as a field of study, a rite of passage to the period of 1942 when the College of Education became the Faculty of Education with M.E. LaZerte as its first dean. Texts for the discipline reflected the ideas of great educators and progressive forms of public schooling with a view to incorporating these standards into the thinking of prospective teachers. The transitional role of such instruction was less evident than it had been in the normal schools and foreshadowed subsequent changes in the history of education at the University.

Until 1945, when the University of Alberta became responsible for all levels of teacher education in the Province, those wanting to become elementary and

intermediate teachers attended normal school. Those aspiring to teach high school enrolled in a concurrent undergraduate program (education and arts and science) or took postdegree studies in the University's School of Education (1929) and, later, the College of Education (1940). As both entities were part of the Faculty of Arts and Sciences, pressure continued from such lobbies as the Alberta Teachers' Alliance, later the Alberta Teachers' Association, which from the mid-1920s held that the field of education deserved the status of a Faculty (see, e.g., Alberta Teachers' Alliance, 1924, p. 17). This happened in 1942 when the University's board of governors agreed with the Department of Education to establish a Faculty of Education. Although this decision did not result in the Faculty being granted sole responsibility for teacher education in the province, it facilitated the setting up of a two-year junior diploma program that entitled its graduates to a grade 7 to 9 teaching certificate (University of Alberta, 1942-1943).

As the diploma program involved accreditation, which formerly had been primarily a normal school responsibility, its course requirements would have interested educational historians and particularly their colleagues in educational philosophy. In this regard they undoubtedly would have been pleased with the attention given their disciplines in the University's Bachelor of Education (BEd) program, especially when compared with measures taken by the Department of Education in the late 1930s to exclude these subjects from the curriculum of the normal schools (Alberta Teachers' Alliance, 1924, pp. 16-17). Diploma students were required to take Philosophy 2 (Introduction to General Psychology and to Logic) or Philosophy 3 (Introduction to Social Psychology and Social Philosophy) from the Faculty of Arts in their first year. Another course in philosophy was required, either Philosophy 51 (History of Philosophy) or Philosophy 54 (Ethics and Social Morality), for students who went on to complete the BEd degree or a joint degree in arts, commerce, or science. With these prerequisites in hand they were permitted to enroll in Education 54 (The Philosophy of Education) a requirement of the final year of the BEd (University of Alberta, 1942-1943).

The 1942 calendar description for Education 54 would have been satisfactory to historians of education who were primarily concerned with studying the ideas of prominent educators and the development of public schooling: "Studies in the Philosophy of Education will be closely associated (1) with a study of the educational classics, and (2) with a survey of modern educational practices in various countries; European, American and Australian" (University of Alberta, 1942-1943, p. 245). That Education 54 kept the same course description given it 30 years before would also have been encouraging to those who believed in the subject's worth. The same parties would likely have been pleased by its place in the final year of the BEd and joint degree programs and by the fact that its status was such that it could not be taken without considerable prior, discipline-related study in the Department of Philosophy.

Syllabuses of the philosophy prerequisites for Education 54 give an idea of the preparation students had in the discipline before taking the latter course. The required text for the History of Philosophy, Philosophy 51, in 1948 was Bernard Russell's *A History of Western Philosophy* (1945),[9] and the class assignment was a 2,500-word essay on "The Contribution of the Ancient Greek

Philosophers to Modern Thought" (Mardiros, 1948). The text in 1951 for Philosophy 54, Ethics and Social Morality, was A.I. Melden's *Ethical Theories: A Book of Readings* (1950) and the assignment was a 2,000-word paper on "J.S. Mill's Utilitarianism" (Mardiros, 1951). As noted above, Education 54 had its roots in the first educational course given at the University of Alberta in the 1912-13 academic year, which was described as an offering in the History of Education. It was renumbered 492 in the 1945-1946 calendar, but it retained its title, the Philosophy of Education, and its earlier description (University of Alberta, 1945-1946, p. 224).

Educational history and philosophy gave increasing attention to progressive education, an orientation that many teacher educators believed greatly enhanced the transitional role traditionally assigned to subjects in these areas. Yet this did not mean the new orientation was always well received. Many of the teachers who responded to a Canada-wide survey in 1948 found much of their professional training to be remote and impractical, or as some put it, "too theoretical [and] time wasted" (Pettifor, 1948, pp. 68-69). And although few specific examples were given in the survey's findings to indicate where this was so in teacher education programs, at least half the respondents quoted singled out the history of education as the chief culprit (Pettifor, 1948, p. 68). The results of this survey were apparently not given much credence at the University of Alberta by such Education 492 instructors as H.E. Smith, who had joined the School of Education in 1929 and was Dean of the Faculty from 1950 to 1955. His 492 course outline in 1951 (Smith, 1951) reflects his adherence (which he wrote about shortly after) to "the philosophic concepts stemming from James, Dewey and Kilpatrick." In Smith's view, the emphasis given by these educators to "the role of intelligence, of experiment, of active learning, of pupil initiative, and the critical use of judgement ... had fairly dominated the [teacher] training schools." Smith's observations are reminiscent of H.C. Newland's enthusiasm for the principles of progressive education. Like Newland, however, Smith was concerned that the principles' "interpretation and application have at times been extravagant and irrational." Yet he had no doubt that the ideas of the progressives "were essentially sound and in keeping with the temper of the times" (Smith, 1956, p. 172).

The content and orientation of the history and philosophy of education offerings from the mid-1930s to the late 1950s tended to be one and the same, even though most were categorized as courses in educational philosophy. Education 594 (formerly Education 104), a graduate seminar in the History of Education at the University of Alberta, was an exception insofar as its title was concerned, but its 1952-1953 calendar description indicates that it was not much different from Education 492, the undergraduate course in educational philosophy:

> History of Education from the Greeks to Modern Times, placing emphasis on great figures and major trends; Comparative—the History of modern education in Canada, Britain, America, France, Germany, Denmark and Russia; and Great Issues in Education—the influence of political, economic, social and religious institutions on education. Theories of the role of education in the modern state. Philosophy of Education. Psychology and Education. (University of Alberta, 1952-1953, p. 282)

That Education 594 cast a wide net is evident from the above prospectus and from a 1948 course syllabus that prescribed no fewer than eight books in modern education and educational history. Although the titles of these required readings ranged from E.W. Knight's *Twenty Centuries of Education* (1940) to A.E. Meyer's *The Development of Education in the Twentieth Century* (1945), the syllabus made it clear that references to educational "thought and practice" in the past would be chosen on the basis of their "contemporary significance" and "of their development or culmination in modern systems of education" (Baker, 1948, p. 1).

Knight's (1940) work was dedicated to Monroe, whose *Text-book in the History of Education* (1905) was the basic reference for Education 54 when it was introduced at the University of Alberta in 1912-1913. *Twenty Centuries* bears a strong resemblance to Monroe's study both in terms of content and ideological orientation. The latter phenomenon is readily apparent in Knight's discussion of the "Value of Educational History" (p. 7), where in keeping with Monroe's views, the subject is assigned its traditional role of providing an "exalted ideal of the teacher's work" during the transitional stage of the rites of passage to teaching. According to Knight, an historical study of the "occupation" of teaching should result in teachers' developing "a desire for higher personal effectiveness" and in heightening their "sense of the dignity and importance of teaching" (p. ix).

Meyer's book (1945, p. ix) on the other hand, inspired in part by the writings of William James and John Dewey, provided prospective teachers with a "bird's-eye glimpse of the recent educational past." Emphasis was placed on progressive education, national systems of education in Europe, and educational change in the United States. Based largely on descriptive analysis, the work does not prescribe one or another approach as the only way of educating the teacher, but offers rather a straightforward discussion of what the neophyte teacher in a wider cultural framework should be expected to know. In this context, state and municipal authorities as the institutional guardians of the rites of passage appear, in Meyer's assessment, "to be favouring more cultural work and less professional work in the training of prospective teachers" (p. 388). The study, then, does not abandon history of education as a rite of passage. As the transition of teacher preparation from the normal schools to the university was taking place, the history of education retained its function as a rite of passage, but with a more contemporary orientation.

Conclusion and Implications

Although there were indications that educational history as a field of study was changing its emphasis by 1950, it continued to provide standards or principles of worth in the transitional stage of the rites of passage to teaching. The subject's place in the Faculty's core curriculum was secure, a security that was based in part on requiring students to have a background in philosophy before taking the compulsory course in the history of education. As was the case in the normal schools in the 1930s, a tension existed between the principal objectives of the course. On the one hand, there remained, as H.M. Tory had argued, a commitment to the liberal education of teachers (Buck, 1993). On the other hand, there was concern about the need to examine the leading values and debates of contemporary education. Along with these objectives, history

of education was expanding to include the study of social questions, particularly relating to the poor, women, and minority groups. The rite of passage to teaching was undergoing change. It was beginning to reflect new directions in educational history that in the coming decades were to reaffirm the discipline's traditional role in the education of teachers.

Acknowledgments

We would like to thank G.H. Buck, B.Y. Card, H.T. Coutts, R.S. Patterson, and B.E. Walker for their observations on the initial draft of this article. Our appreciation is also extended to Sun-Yo Kim for his research assistance and W.H. Swift for his comments on the 1936 Summer School at the University of Alberta. Gratefully acknowledged are Mark Primmer's editorial suggestions, and Joan White and David Kales for their assistance in preparing the manuscript.

Notes

1. The terms *educational history* and *history of education* are considered to be identical.
2. The terms *teacher education* and *teacher training* are used interchangeably in this article.
3. The first reference to the inclusion of educational history in the School Management course occurs in the *Report of the Principal of the Prince of Wales College and the Provincial Normal School* 1908, Appendix E, p. 3. Earlier reports, especially those on School Management examinations, indicates that this practice had been in effect for some time. See, for example, Prince Edward Island, 1888, pp. 105-106, 110; 1891, p. 92.
4. In 1858 the Council of Public Instruction for Canada West approved a program of study for the Toronto Normal School that included the subject of "education." Although the course emphasized matters relating to "School Organization [and] Management," first-class teacher training candidates were also expected to gain "a knowledge of the leading principles of Mental and Moral Philosophy" (*The Normal School of Ontario*, 1871, pp. 11, 13-14, 24-25). lectures on philosophy were linked to practice teaching sessions in the model school (Ontario, *Annual Report Normal Model High and Public Schools*, 1877, p. 98).
5. Graves' book was the required history of education text at all three normal schools in 1931 (Alberta Department of Education, 1931b, p. 11).
6. Newland praised G.K. Haverstock's (principal of the Camrose normal school) course and his use of a text of the same name, *Elementary School Organization and Management* by Dougherty, Gorman, and Phillips (1936) as presenting "a point of view [that fully supports] our new setup."
7. B.Y. Card, who set up a first year history of education course at the University of Alberta in the late 1950s, provided information on the adoption of Counts' text (B.Y. Card, personal interview, 22 June 1992).
8. For a discussion of the relationship between the summer school offerings and the new school program, see "Courses of Special Interest to Teachers at the Present Time" in Alberta Department of Education (1936, pp. 15-17).
9. The volume consisted of three books: Ancient Philosophers, Catholic Philosophy, and Modern Philosophy.

References

Alberta Teacher's Alliance. (1924). Annual general meeting of the Alberta Teacher's Alliance. ATA Magazine, 4(12).

Alberta Department of Education. (1908). *Second annual report of the Department of Education of the Province of Alberta, 1907*. Edmonton: Government Printer.

Alberta Department of Education. (1909). *Third annual report of the Department of Education of the province of Alberta, 1908*. Edmonton: Government Printer.

Alberta Department of Education. (1921). *Fifteenth annual report of the Department of Education of the province of Alberta, 1920*. Edmonton: King's Printer.

Alberta Department of Education. (1922). *Sixteenth annual report of the Department of Education of the province of Alberta, 1921*. Edmonton: King's Printer.

Alberta Department of Education. (1923). *Seventeenth annual report of the Department of Education of the province of Alberta, 1922*. Edmonton: King's Printer.

Alberta Department of Education. (1924). *Eighteenth annual report of the Department of Education of the province of Alberta, 1923*. Edmonton: King's Printer.

Alberta Department of Education. (1925). *Nineteenth annual report of the Department of Education of the province of Alberta, 1924*. Edmonton: King's Printer.
Alberta Department of Education. (1926. *Twentieth annual report of the Department of Education of the province of Alberta, 1925*. Edmonton: Acting King's Printer.
Alberta Department of Education. (1927). *Twenty-first annual report of the Department of Education of the province of Alberta, 1926*. Edmonton: Acting King's Printer.
Alberta Department of Education. (1928). *Twenty-second annual report of the Department of Education of the province of Alberta, 1927*. Edmonton: King's Printer.
Alberta Department of Education. (1929). *Twenty-third annual report of the Department of Education of the province of Alberta, 1928*. Edmonton: King's Printer.
Alberta Department of Education. (1930). *Twenty-fourth annual report of the Department of Education of the province of Alberta, 1929*. Edmonton: King's Printer.
Alberta Department of Education. (1931a). *Twenty-fifth annual report of the Department of Education of the province of Alberta, 1930*. Edmonton: King's Printer.
Alberta Department of Education. (1931b). *Alberta normal schools announcement*. Edmonton: King's Printer.
Alberta Department of Education. (1932). *Twenty-sixth annual report of the Department of Education of the province of Alberta, 1931*. Edmonton: King's Printer.
Alberta Department of Education. (1933. *Twenty-seventh annual report of the Department of Education of the province of Alberta, 1932*. Edmonton: King's Printer.
Alberta Department of Education. (1934). *Twenty-eighth annual report of the Department of Education of the province of Alberta, 1933*. Edmonton: King's Printer.
Alberta Department of Education. (1935). *Twenty-ninth annual report of the Department of Education of the province of Alberta, 1934*. Edmonton: King's Printer.
Alberta Department of Education. (1936). *Joint summer session 1936*. Edmonton: Department of Education.
Bailyn, B. (1960). *Education in the forming of American society*. New York: Random House.
Baker, H.S. (1948). *Education 594, History of education summer session syllabus, 1948*. University of Alberta Archives Acc. 69-33-38.
Battersby, W. (Ed.). (1952). *De la Salle: Letters and documents*. London: Longmans Green.
Black, W.G. (1936). *The development and present status of teacher education in Western Canada, with special reference to the curriculum*. Unpublished doctoral dissertation, University of Chicago.
Browning, O. (1881). *An introduction to the history of educational theories*. London: Paul, Trench.
Buck, G.H. (1993). Faculty of Education, University of Alberta: A 30-year gestation. *Alberta Journal of Educational Research, 39*, 319-334.
Carr, E.H. (1987). *What is history?* Harmondsworth, UK: Penguin.
Chapple, E.D., & Coon, C.C. (1942). *Principles of anthropology*. New York: Holt.
Counts, G. (1932). *Dare the school build a new social order?* New York: Arno Press.
Cremin, L. (1955). The recent development of the history of education as a field of study in the United States. *History of Education Journal, 7*(1), 1-35.
Davidson, T. (1900). *A history of education*. New York: Appleton.
Donohue, J. (1963). *Jesuit education: An essay on the foundations of an idea*. New York: Fordham University Press.
Dougherty, J.H., Gorman, F.H., & Phillips, C.A. (1936). *Elementary School Organization and Management*. New York: Macmillan.
Eby, F. (1952). *The development of modern education*. Englewood Cliffs, NJ: Prentice-Hall.
Eisenhart, M., Behm, L. & Romagnano, L. (1991) Learning to teach: Developing expertise or rite of passage? *Journal of Education for Teaching, 17*(1), 51-69.
Farrell, A. (1938). *The Jesuit code of liberal education: Development and scope of the Ratio Studiorum*. Milwaukee, MN: Bruce.
Ferruolo, S. (1985). *The origins of the university: The schools of Paris and their critics*. Stanford, CA: Stanford University Press.
Forrester, A. (1867). *The teacher's text book*. Halifax, NS: Mackinaly.
Fuch, E. (1969). *Teachers talk: Views from inside city schools*. Garden City, NY: Doubleday.
Gates, C.E. (1989). Image, imagination, and initiation: Teaching as a rite of passage in the novels of L.M. Montgomery and Laura Ingalls Wilder. *Children's Literature in Education, 20*(3) 165-173.
Gauss, G.F. (1990). *Value and justification: The foundations of liberal theory*. Cambridge, MA: Cambridge University Press.

Gehrke, N.J. (1991). Seeing our way to better helping of beginning teachers. *Educational Forum, 55*(3), 233-242.
Graves, W.G. (1925). *A student's history of education.* New York: Macmillan.
Hale, L., & Staratt, R.J. (1989). Rites of passage: A case study of teacher preparation. *Journal of Educational Administration, 27*(3), 24-29.
Harvey, J. (1986). The Church in Canada twenty years after Vatican II. *Lumen vitae, 41*(3), 282-290.
Hewson, C.E. (1940). *The history of commercial education in Canada.* Unpublished master's thesis, University of Alberta.
Hodysh, H.W. (1980). Research traditions in the history of education: A study of goals and their emergence. *Alberta Journal of Educational Research, 26,* 51-59.
Horne, H.H. (1905). *The philosophy of education: Being the foundations of education in the related natural and mental sciences.* New York: Macmillan.
Houston, S.E., & Prentice, A. (1988). *Schooling and scholars in nineteenth-century Ontario.* Toronto: University of Toronto Press.
Joint Committee. (1929). *Minutes.* Edmonton: University of Alberta Archives.
Kaplan, A. (1964). *The conduct of inquiry.* San Francisco: Chandler.
Kimball, B. (1986). *Orators and philosophers: A history of the idea of liberal education.* New York: Teachers College, Columbia University.
Kimball, B. (1992). *The "True professional ideal" in America.* Cambridge, MA: Blackwell.
Knight, E.W. (1940). *Twenty centuries of education.* Boston: Ginn.
Koos, L., & Woody, G. (1919). The training of teachers in the accredited high schools in the state of Washington. In *Eighteenth Yearbook of the National Society for the Study of Education* (pp. 246-257. Bloomington, IL: Public School Publishing.
Kuhn, T.S. (1977). Objectivity, value judgments and theory choice. In T.S. Kuhn (Ed.), *The essential tension* (pp. 320-337). Chicago: Chicago University Press.
Lacey, A.C. (1977). *The socialization of teachers.* London: Methuen.
Lacey, A.C. (1987). Professional socialization of teachers. In M. J. Donkin (Ed.), *International encyclopedia of teaching and teacher education.* Oxford: Pergamon.
Le Goff, J. (1980). *Time, work and culture in the middle ages.* Chicago: University of Chicago Press.
Mann, G. (1961). *Alberta normal schools: A Descriptive study of their development.* Unpublished master's thesis, University of Alberta.
Mardiros, A.M. (1948). *Summer session syllabus 1948, Philosophy 51.* University of Alberta Archives, Acc. 69-33-39.
Mardiros, A.M. (1951). *Syllabus for philosophy 54, summer session 1951.* University of Alberta Archives Acc. 69-33-39.
Melden, A.I. (1950). *Ethical theories: A book of readings.* New York: Prentice-Hall.
Meyer, A.E. (1945). *The development of education in the twentieth century.* New York: Prentice-Hall.
Miklos, E., Conklin, R., & Greene, M. (1987). *Teachers' evaluation of their preparation for teaching.* Edmonton, AB: Alberta Teachers' Association.
Monroe, P. (1905). *A text-book in the history of education.* New York: Macmillan.
Montgomery, L.M. (1981). *Anne of Avonlea.* Toronto: McClelland and Stewart-Bantam.
Montgomery, L.M. (1985). *The selected journals of L.M. Montgomery Vol. 1, 1889-1910* (M. Rubio & E. Watterston, Eds.). Toronto: Oxford University Press.
Myrehaug, D. (1972). *M.E. LaZerte: Contributions to teacher education in Alberta.* Unpublished master's thesis, University of Alberta.
Newland, H.C. (1936). *The training and certification of teachers of Alberta's schools.* Edmonton: Project Yesteryear, Faculty of Education, University of Alberta.
Normal school of Ontario: Its design and functions, chiefly taken from the report of the chief superintendent of education for the year 1869. Toronto: Hunter, Row.
North-West Territories. (1897). *Report of the Council of Public Instruction of the North-West Territories of Canada, 1896.* Regina: Queen's Printer.
North-West Territories. (1898). *Report of the Council of Public Instruction of the North-West Territories of Canada, 1898.* Regina: Queen's Printer.
North-West Territories. (1903). *Annual report of the Department of Education of the North-West Territories.* Regina: Queen's Printer.
Ontario. (1877). *Annual report normal model high and public schools (1875).* Sessional Papers, 1877(31).
Ontario. (1882). *Report of the department of education.* Sessional Papers, 1882(5).
Ontario. (1891). *Sessional papers.* First Session of Seventh Legislature of the Province of Ontario, Vol. XXIII Part II.

Ottewell, A.E. (1915). *The university extension movement.* Unpublished master's thesis, University of Alberta.
Painter, F.V.N. (1901). A history of education. New York: Appleton.
Patterson, R.S. (1974). Hubert C. Newland: Theorist of progressive education. In R.S. Patterson, J.W. Chalmers, & J.W. Friesen (Eds.), *Profiles of Canadian educators.* Toronto: Heath.
Pettifor, R.E. (1948). *Public opinion concerning the selection and training of teachers.* Unpublished master's thesis, University of Alberta.
Phillips, C.E. (1957). *The development of education in Canada.* Toronto: Gage.
Prince Edward Island. (1888). *Annual report of the public schools of Prince Edward Island for the year 1888.* Charlottetown: King's Printer.
Prince Edward Island. (1891). *Annual report of the public schools of Prince Edward Island for the year 1891.* Charlottetown: King's Printer.
Quick, R. (1895). *Essays on educational reformers.* New York: Appleton.
Rashdall, H. (1936). *The universities of Europe in the middle ages, Vol. 1.* London: Oxford University Press.
Report of the Committee on Secondary School Studies with reports of the conferences arranged by the Committee. (1894). New York: American Book Company.
Report of the principal of Prince of Wales College and the Provincial Normal School. (1908). Annual Report of the public schools of Prince Edward Island, Appendix E.
Rosenkranz, J.K. (1890). *The philosophy of education.* New York: Appleton.
Russell, B. (1945). *A history of western philosophy.* New York: Simon & Schuster.
Schwickerath, R. (1903). *Jesuit education.* St. Louis, MO: Herder.
Smith, H.E. (1951). *Education 492, summer session 1951.* University of Alberta Archives Ac. 69-33-41.
Smith, H.E. (1956). Teacher training. In J. Katz (Ed.), *Canadian education today.* Toronto: McGraw-Hill.
Spencer, H. (1911). *Essays on education and kindred subjects.* London: Dent.
Stamp, R.E. (1982). *The schools of Ontario, 1876-1976.* Toronto: University of Toronto Press.
Sylwester, R. (1987). Rites of passage: An elementary curriculum course that focuses on the student to teacher transition. *Teaching Education, 1*(1), 29-32.
Thomas, C. (1969). *Ryerson of Upper Canada.* Toronto: Ryerson Press.
Thomson, K.H. (1938). *The educational philosophy of Tolstoy.* Unpublished master's thesis, University of Alberta.
Thomson, K.H. (1939). *The educational philosophy of Robert Owen.* Unpublished graduate thesis, University of Alberta.
Tinto, V. (1988). Stages of student departure. *Journal of Higher Education, 59*(4), 438-455.
Tompkins, A. (1894). *Philosophy of teaching.* Boston: Ginn.
University of Alberta. (1912-1913). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1913-1914). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1914-1915). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1915-1916). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1929-1930). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1935-1936). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1936-1937). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1940-1941). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1942-1943). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1945-1946). *Calendar.* Edmonton: University of Alberta.
University of Alberta. (1952-1953). *Calendar.* Edmonton: University of Alberta.
University of Alberta Senate. (1911). *Minutes.* Edmonton: University of Alberta Archives.
van Gennep, A. (1960). *The rites of passage.* London: Routledge and Kegan Paul.
White, J.J. (1989). Student teaching as a rite of passage. *Anthropology and Education Quarterly, 20,* 177-195.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 57-65*

Eileen Wood
Wilfrid Laurier University

Teena Willoughby

Violet Kaspar
University of Waterloo

and

Tracey Idle
Wilfrid Laurier University

Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations

The purpose of this study was to examine the potency of elaborative interrogation relative to supported learning contexts for adolescent learners. One hundred and twenty students were instructed to use one of four learning strategies to learn animal facts: a questioning strategy that employs why *questions (elaborative interrogation); a judgment condition where students assessed the quality of provided elaborations; and two repetition conditions where students either repeated experimenter-provided elaborations or non-elaborated facts. As in existing research with young children and university students, these 10th- to 12th-graders demonstrated greater learning when instructed to use elaborative interrogation relative to repetition of elaborated or non-elaborated facts. Although elaborative interrogation did not significantly differ from the judgment condition, students indicated that they accessed prior knowledge more often when employing the elaborative interrogation strategy, suggesting that elaborative interrogation facilitates learning by encouraging access to prior knowledge.*

Le but de cette étude était d'examiner l'efficacité de l'interrogation élaborative comparée à des contextes d'apprentissage à l'appui pour les adolescent(e)s apprenant(e)s. On a donné des instructions à cent vingt élèves du niveau secondaire à utiliser une de quatre stratégies d'apprentissage afin d'apprendre des faits sur les animaux: une stratégie de questionnement qui utilise surtout des questions de "pourquoi" (interrogation élaborative); une situation de jugement qui obligeait les élèves à évaluer la qualité des faits et des élaborations présentés; et deux conditions de répétition qui contraignaient les élèves à soit répéter les faits présentés par les expérimenteurs ou bien à répéter des faits non élaborés. Ces apprenant(e)s de la 10^{e} à la 12^{e} année ont démontré qu'ils apprenaient plus lorsqu'ils utilisaient la méthode de l'interrogation élaborative plutôt que celle de la répétition de faits élaborés et des faits non élaborés. Ceci se compare bien aux recherches qui existent faites sur les jeunes apprenant(e)s et les étudiant(e)s de l'université. Même s'il y a peu de différences entre l'interrogation élaborative

Eileen Wood is an assistant professor.
Teena Willoughby holds a postdoctoral position.
Violet Kaspar is enrolled in the doctoral program.
Tracey Idle has just completed her master's degree.
All authors have designated interests in developmental psychology with specific interest in cognitive development and the development of memory strategies.

et la situation de jugement, les élèves ont indiqué qu'ils(elles) se servaient des connaissances apprises antérieurement plus souvent lorsqu'ils(elles) utilisaient les stratégies de l'interrogation élaborative. Ceci suggère que l'interrogation élaborative facilite l'apprentissage puisqu'elle encourage l'élève à avoir recours aux connaissances apprises antérieurement.

Recent research has advocated using an associative learning strategy called elaborative interrogation to facilitate learning from prose (Pressley, Symons, McDaniel, Snyder, & Turnure, 1988; Woloshyn, Willoughby, Wood, & Pressley, 1990; Wood, Pressley, Turnure, & Walton, 1987). Elaborative interrogation encourages learners to integrate to-be-learned information with existing knowledge by requiring learners to respond to *why* questions (i.e., Why would this fact be true?). Existing research has examined elaborative interrogation with very young learners (e.g., Wood, Miller, Symons, Canough, & Yedlicka, in press), fourth through eighth graders (Wood, Pressley, & Winne, 1990), and university students (e.g., Pressley, McDaniel, Turnure, Wood, & Ahmad, 1987). Performance differences across these studies suggest that the ability to effectively use elaborative interrogation increases with advancing age (Wood et al., 1990). The present study completes the existing developmental picture for this strategy by including an adolescent population. The potency of elaborative interrogation is also compared with supported learning contexts where elaborations are provided.

In a series studies, Bransford et al. (1982; Franks et al., 1982; Stein & Bransford, 1979; Stein et al., 1982) demonstrated greater learning of factual information for 5th graders and adults when they studied sentences that were accompanied with elaborations versus non-elaborated sentences. In addition, elaborations that clarified the conceptual relations within the facts produced greater learning than less explanatory elaborations. For example, memory for the sentence "The tall man bought the crackers" was higher when the sentence was followed by the elaboration "that were on the top shelf" than the elaboration "that were on sale." Bransford et al. argued that the former elaboration was more powerful because it specifically explained the relation between the type of man and the activity. The explanation reduced the arbitrariness of the relations and hence made the material more meaningful and more memorable for the learner.

Pressley et al. (1987), however, found that adults experienced these performance gains only when they were unaware of an upcoming memory test while studying (incidental learning contexts). When adult learners were told to expect a memory test (Intentional learning), as were all the children in the Bransford et al. studies (1982), there was no advantage for providing elaborations relative to simple repetition of the non-elaborated sentences (Pressley et al., 1987). The lack of differences between these two conditions most probably reflects the tendency for adult learners in the non-elaborated repetition condition to spontaneously elaborate some of the sentences when studying intentionally. Alternatively, some of the provided elaborations may have conflicted with the learners' existing knowledge, resulting in interference and deflated performance.

In contrast, Pressley et al. (1987) found robust memory gains for adults who were instructed to generate their own elaborations. These students were asked to answer *why* questions. Adults taught to use the elaborative interrogation

strategy exceeded the performance of adults who rehearsed either elaborated or non-elaborated sentences. Elaborative interrogation was thought to prompt learners to make the new information more meaningful by relating the new information to the learners' existing prior knowledge (Willoughby, Waller, Wood, & MacKinnon, 1993).

Wood et al. (1990) extended this research to child populations. Using the same *man* sentences as Bransford et al. (1982), they found that 4th- to 8th-graders using elaborative interrogation also outperformed those who repeated elaborated and non-elaborated sentences. Consistent with the early Bransford et al. (1982) studies, the older grade school children in this study also seemed to benefit from the support offered by provided elaborations relative to non-elaborated sentences. These younger learners presumably have less developed knowledge bases and are less likely to spontaneously elaborate information than their adult peers. However, they can be compensated to some extent by providing elaborations that clarify and reduce the arbitrariness of new information (Wood et al., 1990). The present study addresses the relative benefits of providing versus self-generating elaborations with adolescents to determine whether provided elaborations support learning of factual information for this population.

In the existing literature (e.g., Wood et al., 1990), provided elaborations are accompanied by instructions to rehearse the information. It is assumed that the presentation of the elaboration will be sufficient to enhance learning. Repetition is a familiar and comfortable strategy for learners of all ages and often serves as the default strategy (Garner, 1990). Although this supported repetition condition (i.e., provided elaboration condition) has both practical and theoretical relevance in its present form, the rehearsal instructions fail to maximize the learning potential of provided elaborations. For example, less cognitive effort (Jacoby, 1978; Tyler, Hertel, McCallum, & Ellis, 1979) is exerted when repeating elaborations than when generating elaborations. In response to this problem, the present study introduced a condition that maximized students' processing of material while supporting their knowledge base through supplied elaborations. In this condition, students were required to judge the adequacy of provided elaborations and explain their judgments. Such a strategy entails active evaluation and reflection while processing the new information, yet poses fewer demands on the learners' prior knowledge. It was expected that this strategy might produce similar learning gains to elaborative interrogation and exceed the performance of simply repeating the provided elaborations. The present study examines the relative efficacy of memory strategies when they are provided versus self-generated, and extends the developmental and experimental research by including an adolescent population.

Method

Subjects

One hundred and twenty high school students (53 male and 67 female) volunteered to participate in this study. The sample comprised 17 10th-graders, 34 11th-graders, and 69 12th-graders. Students were drawn from two schools in Southern Ontario and ranged in age from 14 to 19 years (M=16 years, 7 months, SD=1 year). Students were randomly assigned to one of four experimental conditions: elaborative interrogation, judgment of provided elaborations, ex-

perimenter-provided elaborations, and repetition. Approximately equal proportions of males and females were assigned to each condition.

Materials and Procedure

Students were tested individually in their school. The session began by training students to use their assigned strategy for three practice sentences. Students were then asked to use that strategy to learn 54 animal facts. After study, students were given a five-item distractor task before completing a 54-item memory test.

All students studied six facts about nine animals plus three practice sentences. The sentences contained information about the living environment, diet, sleep habits, predators, preferred and global habitat for each of the following nine animals: Grey Seal, Townsend Mole, Emperor Penguin, Little Brown Bat, Blue Whale, House Mouse, Swift Fox, Western Spotted Skunk, and American Pika. Factual content of the sentences was verified for both 4th- to 8th-graders and adults (see Wood et al., 1990; Willoughby et al., 1993, respectively). These materials presented information that was novel, yet drew from a topic domain of which learners had some general knowledge.

Four sets of study sentences were constructed, one for each of the elaborative interrogation and repetition conditions and two sets for the provided conditions. Sentences were presented one at a time with an accompanying study prompt. In the elaborative interrogation condition, each set of sentences was followed by the study prompt "Why would that animal do/have that?" In the repetition condition, each fact was followed by the prompt "Read the sentence at a rate that allows you to understand that the fact is true." For the provided elaboration conditions sentence extensions accompanied each sentence. An adequate and an inadequate elaboration were constructed for each fact. Adequate elaborations explained the significance of the animal engaging in the specific behavior (e.g., *The grey seal lives on exposed rocky coasts* so that the sun can warm the rocks before it lies on them.) while inadequate elaborations failed to clarify why that animal in particular would engage in the described behavior (e.g., *The grey seal lives on exposed rocky coasts* because it lives near water.). The two types of elaborations were included to ensure that students in the judgment condition perceived their task as both interesting and valuable. Half of the adequate elaborations for each animal were assigned to one set and the remaining half to the other set. Likewise, half the inadequate items were assigned to each set. Therefore, the two sets of animal facts were counterbalanced for the presentation of adequate and inadequate elaborations in the provided elaboration conditions.

Students in the elaborative interrogation condition generated responses to the *why* question. The experimenter specified that good answers to *why* questions explain why a given fact is true of that animal in particular and no other animal. During the practice trials, the experimenter provided feedback and prompting until an adequate elaboration was generated. Students in the repetition and experimenter-provided elaboration condition were instructed to repeat each sentence aloud at a rate that would enable them to comprehend the sentence so that they could recall the information later. Students in the provided elaboration condition were also told that understanding the provided elaboration would help them to remember the information. Students who

judged the provided elaborations were instructed to evaluate whether the elaboration was "good" or "poor" and to justify their evaluation. They were told that a good elaboration (adequate) would explain why only that particular animal rather than any other animal would engage in the activity and that poor elaborations (inadequate) would fail to clarify this relation.

After studying the three practice sentences, students in all four conditions completed a sample recall test. Students were presented the animal behavior in the form of a *which* question and their task was to generate the animal name. For example, "Which animal lives on exposed rocky coasts?" (Answer: grey seal).

After the practice session, students were reminded of their strategy instructions. Students were then presented with the 54 to-be-learned animal facts that were prerecorded on audiotape with 18-second intervals between each sentence. This was done to ensure that all students were exposed to the complete fact and to minimize difficulties resulting from reading ability. The interval following each sentence provided time for the learners to employ their study strategy. All students were instructed to study aloud and their responses were recorded on audiotape. This provided a means to ensure that students were using the assigned strategy at study. No feedback was provided during the study phase.

Prior to the memory test all students completed a five-item distractor task that asked questions about students' general interest in animals. This task was included to reduce recency effects. Memory was then measured via a 54-item cued recall task. Students were provided with the animal behavior and asked to generate the appropriate animal name. Recall test items were asked in a different random order for each student.

Results

The main analyses were conducted on the memory test data. Memory for the animal facts was compared across the four study conditions using an analysis of variance (ANOVA). The mean recall scores are reported in Table 1. There was a significant effect for study condition $F(3, 116)=6.28$, $p<.001$. Differences were assessed using Tukey's HSD post hoc comparison procedure. Consistent with our expectation, recall performance in the elaborative interrogation condition did not differ from the judgment condition, $t(116)=2.20$, $p>.05$. Students in the elaborative interrogation condition significantly outperformed students in both the repetition and the experimenter-provided elaboration condition,

Table 1
Mean Recall Scores Across Experimental Conditions

	Mean	*SD*
Elaborative Interrogation	34.13	6.99
Judgment	29.33	8.83
Provided Elaboration	26.53	9.32
Repetition of Facts of Non-elaborated Facts	25.50	8.43

Note. Maximum score for each group is 54. $n=30$ per cell.

$t(116)=3.96$, $p<.05$ and 3.49, $p<.05$ respectively, consistent with previous research. No other comparisons were significant, largest $t(116)=1.76$, $p>.05$ for the judgment versus repetition comparison.

Relation of Quality of Responses to Memory Performance

Elaborative interrogation. Study responses were examined to assess the impact of the quality of response generated at study on subsequent recall. The elaborations were scored as adequate, inadequate, or no response. Adequate responses clearly specified the significance of the behavior for that particular animal. Inadequate elaborations included simple restatements, incomplete responses, or explanations that failed to clarify why that animal would engage in that behavior. Adequate responses were further categorized into correct, incorrect, and pat categories. These categories were constructed to reflect the source of information that the students accessed when answering the *why* questions in order to determine whether access to appropriate (i.e., scientifically correct) information was necessary for the success of this strategy or whether access to less accurate or less specific preexisting knowledge would promote learning. Correct responses corresponded with expert knowledge of the animal. Incorrect elaborations used information that was not true of the animal. Pat responses were elaborations that were general or vague.

Interrater reliability was established by two raters who scored over 30% of the elaborations with 95% agreement. Differences were resolved by discussion. The remaining data were scored by one of the two raters.

Overall, the majority of elaborations were adequate (60%). Students generated inadequate elaborations for 33% of the facts and failed to generate a response for only 7% of the facts (see Table 2 for mean scores). Of the adequate responses, the majority of elaborations were factually correct (74%).

A series of item-by-item conditional probabilities was calculated to determine the relation between the quality of responses provided at study and subsequent memory performance. Each elaboration provided at study was matched to the corresponding response on the memory test (means are reported in Table 2). Two sets of comparisons were conducted using Tukey's HSD procedure, one set for the adequate, inadequate, and no response categories and one for the correct, incorrect, and pat distinctions. The probability of

Table 2
Frequencies and Probabilities of Correct Recall as a Function of Adequacy of Response in the Elaborative Interrogation Condition

	Conditional Probabilities			*Frequencies*		
	n	*Mean*	*SD*	*n*	*Mean*	*SD*
No Response	19	.42	.29	30	3.97	6.33
Inadequate	30	.62	.18	30	17.53	7.95
Adequate	30	.67	.12	30	32.50	10.15
Correct Adequate	30	.71	.11	30	23.93	8.77
Incorrect Adequate	29	.52	.39	30	3.47	2.81
Pat Adequate	30	.57	.29	30	5.10	2.88

recalling an item was greater if learners generated an adequate or an inadequate response than if they failed to respond, $t(76)=4.39$, $p<.05$ and $t(76)=3.51$, $p<.05$, respectively. Memory performance did not differ following the generation of adequate versus inadequate elaborations, $t(76)=1.00$, $p>.05$.

The probability of correct recall was higher for correct elaborations than incorrect elaborations, $t(86)=2.57$, $p<.05$. There were no other significant differences, $t(86)=1.89$, $p>.05$ for the correct versus pat comparison and $t(86)=.68$, $p>.05$ for the incorrect versus pat comparison.

Discussion

Consistent with previous research involving children (Wood et al., 1990) and adults (Pressley et al., 1987), this study demonstrated that adolescents remember more when they use elaborative interrogation relative to rehearsal of elaborated and non-elaborated material. As in the adult population, provided elaborations did not offer any advantage for these students relative to simple repetition of the non-elaborated materials. There were no differences between the elaborative interrogation and judgment conditions, suggesting that there is some advantage for supporting students' learning by providing elaborations, but only when students are encouraged to actively process the information in a meaningful fashion.

In our judgment condition, students were encouraged to be more active by explicit directions to use the provided elaborations to understand the to-be-learned material. Requiring students to evaluate the provided elaborations encouraged them to determine whether the elaboration would help them to understand the information, which in turn made them focus on the relations presented in the to-be-learned material. It was expected that the cognitive effort and hence the processing expended in the elaborative interrogation and judgment conditions would be comparable. This was supported by the lack of significant differences in memory performance between the elaborative interrogation and judgment conditions. However, it could be argued that elaborative interrogation may provide some advantage relative to the judgment condition. For example, the judgment condition did not outperform the provided elaboration or repetition condition as did elaborative interrogation.

This possible advantage for the elaborative interrogation strategy may be attributed to the way students process information when using this strategy. For example, elaborative interrogation may promote greater integration of the material. Integration may be seen at two levels. First, students are encouraged to draw from their own prior knowledge to respond to the question, thereby integrating the to-be-learned information within existing schemata (Willoughby, Wood, & Khan, in press). Second, students may be better able to understand the information if they integrate the facts as they are presented in the sequence. Students who use information provided early in the animal set to understand later facts may develop stronger interconnections and associations among the facts that enhances both understanding and retrieval.

Instructions to judge the provided elaborations, in contrast, may encourage students to actively associate the items within a fact but not process the facts holistically for each animal. In fact, we examined this possibility by contrasting both the number of times students generated information that clearly represented prior knowledge and the number of times they integrated previous facts

given during study. Elaborative interrogation students were more likely to make reference to existing prior knowledge and to previous presented facts when studying than students in the judgment condition, largest $t(57)=11.07$, $p<.05$ for the prior knowledge contrast. This suggests that we need to encourage students to go beyond processing discrete or independent items, perhaps by providing explicit instructions to access their prior knowledge.

Performance in the provided elaboration condition was not greater than that in the simple repetition condition. This parallels similar findings in the adult literature when students are exposed to an intentional learning situation (Pressley et al., 1987). Adolescents, like adults, may be able to mediate performance through the spontaneous activation of more complex strategies when instructed to rehearse non-elaborated facts, hence elevating performance. Alternatively, these students may have been distracted by the provided elaborations to the extent that they were unable to understand when the arbitrariness of the to-be-learned information was reduced (Pressley, Wood, & Woloshyn, 1990). In either case, simply providing students with information is insufficient to promote learning. Adolescents, like most learners, require explicit instructions to actively and thoroughly process to-be-learned material.

The findings in this study also are consistent with the trend observed by Wood et al. (1990) with grade school children. That is, generating a response to the *why* question, whether adequate or inadequate, resulted in greater probability of correct recall than if a response was not made at all. When younger students fail to generate an elaboration, it might be that they are not processing the new information thoroughly, and this partial processing may not be sufficient to allow the information to be embedded in their knowledge base. Adults, on the other hand, demonstrate equivalent performance across these categories, suggesting that the mere attempt to generate a response activates a network of information related to the to-be-learned facts (Pressley et al., 1987; Woloshyn et al., 1990). In the present study, the probability of correct recall was much higher following generation of a correct elaboration than an incorrect one. This pattern appears to be consistent across age (Willoughby et al., 1993; Wood et al., 1990) and verifies the importance of access to an appropriate knowledge base.

In summary, the pattern of performance in this study with adolescents most closely mirrors that of grade school children both when comparing across study strategies and the impact of quality in the elaborative interrogation condition. Adding an active component, as in the judgment condition, provides some added support for these students, but their performance is maximized when they are explicitly instructed to use their prior knowledge through strategies such as elaborative interrogation.

Acknowledgments

Participation of the first and third authors was made possible through grants from the Canadian Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council, respectively.

Note

Correspondence about this article should be directed to Eileen Wood, Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5.

References

Bransford, J.D., Stein, B.S., Vye, N.J., Franks, J.J., Auble, P.M., Mezynski, J.J., & Perfetto, G.A. (1982). Differences in approaches to learning: An overview. *Journal of Experimental Psychology: General, 111*, 390-398.

Franks, J.J., Vye, N.J., Auble, P.M., Mezynski, K.J., Perfetto, G.A., Bransford, J.D., Stein, B.S., & Littlefield, J. (1982). Learning from explicit versus implicit texts. *Journal of Experimental Psychology: General, 111*, 414-422.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research, 60*, 517-529.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649-667.

Pressley, M., McDaniel, M.A., Turnure, J.E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 291-300.

Pressley, M., Symons, S., McDaniel, M.A., Snyder, B.L., & Turnure, J.E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology, 80*, 268-278.

Pressley, M., Wood, E., & Woloshyn, V.E. (1990). In W. Schneider & R.E. Weinert (Eds.), *Interactions among aptitudes, strategies, and knowledge in cognitive performance* (pp. 200-221). New York: Springer-Verlag.

Stein, B.S., & Bransford, J.D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*, 769-777.

Stein, B.S., Bransford, J.D., Franks, J.J., Owings, R.A., Vye, N.J., & McGraw, W. (1982). Differences in the precision of self-generated elaborations. *Journal of Experimental Psychology: General, 111*, 399-405.

Tyler, S.W., Hertel, P.T., McCallum, M.C., & Ellis, H.C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 607-617.

Willoughby, T., Waller, T.G., Wood, E., & MacKinnon, G.E. (1993). The effect of prior knowledge and prose structure on immediate and delayed recall following training in elaboration. *Contemporary Educational Psychology, 18*(36), 36-46.

Willoughby, T., Wood, E., Khan, M. (1993). Isolating variables that impact or detract from the effectiveness of elaboration strategies. *Journal of Educational Psychology, 94*, 245-254.

Woloshyn, V.E., Willoughby, T., Wood, E., & Pressley, M. (1990). Elaborative interrogation facilitates adult learning of factual paragraphs. *Journal of Educational Psychology, 82*, 513-524.

Wood, E., Miller, G., Symons, S., Canough, T., & Yedlicka, J. (in press). Effects of elaborative interrogation on the young learner's recall of fact. *Elementary School Journal.*

Wood, E., Pressley, M., Turnure, J.E., & Walton, R. (1987). Enriching children's recall of picture-dictionary definitions with interrogation and elaborated pictures. *Educational Communication and Technology Journal, 35*(1), 43-52.

Wood, E., Pressley, M., & Winne, P.H. (1990). Elaborative interrogation effects on children's learning of factual content. *Journal of Educational Psychology, 82*, 741-748.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 67-94*

Joseph Blase

and

Jo Roberts
University of Georgia

The Micropolitics of Teacher Work Involvement: Effective Principals' Impacts on Teachers

Although the micropolitics of educational settings has received significant attention in recent years, few empirical studies of this important phenomenon have been published. The data discussed in this article were drawn from a larger, open-ended qualitative study that examined the perspectives of 1,200 teachers on the everyday political strategies school principals use to influence them. Because of their theoretical significance, the responses of the 836 teachers who specifically described their principals as open and effective are the focus of this article. The data strongly suggest that the strategic orientation of effective principals—here described in terms of the emergent concept of normative-instrumental leadership—has a positive impact on teacher work involvement. Such involvement, the data argue, develops primarily because the influence strategies and related goals pursued by open and effective principals are congruent with teachers' moral/value dispositions. A second concept derived from the data—normative-instrumental involvement—outlines the teachers' specific affective, cognitive, and behavioral responses to effective principals. Theoretical ideas relevant to this concept are briefly discussed using elements of Etzioni's (1961, 1975) compliance theory (Glaser, 1978; Glaser & Strauss, 1967). The study's implications for school restructuring and teacher empowerment are also explored.

Quoique la micropolitique dans le milieu éducationnel s'est mérité une certaine attention depuis quelques années, très peu d'études empiriques ont été publiées sur ce phénomène important. Les données discutées dans cet article ont été tirées d'une plus grande étude qualitative qui cherchait à sonder les perspectives de 1200 enseignant(e)s sur les stratégies politiques quotidiennes qu'utilisaient les directeurs et les directrices d'école pour influencer le corps professoral. Cet article examine de près les réponses de 836 enseignant(e)s qui ont décrit leurs directeurs et directrices d'école comme étant particulièrement ouvert(e)s et efficaces. Ces réponses ont une valeur théorique importante. Les données suggèrent fortement que l'orientation stratégique des directeurs et directrices efficaces–étant décrite en terme d'un concept émergent d'un leadership normatif-instrumental–auraient un impact positif sur le niveau de participation des enseignant(e)s dans leur milieu de travail. Les données énoncent que cette participation dans le milieu de travail croît surtout parce que les stratégies d'influence et les buts reliés poursuivis par les directeurs et directrices d'école ouvert(e)s et efficaces correspondent de façon directe avec les principes moralité/valeur des enseignant(e)s. De ces données provient un deuxième concept–celui de la participation normative-instrumentale–qui décrit les réactions spécifiques affectives, cognitives, et comportementales des enseignant(e)s envers les directeurs et directrices efficaces. En se servant des éléments de la théorie d'acquiescement d'Etzioni (1961, 1975), on discute brièvement quelques idées théori-

Joseph Blase is a professor of educational leadership. Current research interests include the micropolitics of education and the sociology of teaching.
Jo Roberts is an associate professor of educational leadership. Her current research focuses on the principalship and supervisory discourse. She has extensive experience as a teacher, principal, director of staff development, and educational consultant.

ques reliées à ce concept (Glaser, 1978; Glaser & Strauss, 1967). On explore également les implications de la restructuration des écoles et de la prise de plein pouvoir des enseignant(e)s.

Over the last few years micropolitical research in education has expanded rapidly. In general, such research focuses on the political-influence strategies individuals and groups use in organizational settings to influence and to protect themselves from others (Ball, 1987; Blase, 1991; Pfeffer, 1981). Ball (1987) described the micropolitics of change, career, gender, and interpersonal relationships, among other things, in British comprehensive schools. From studies of teachers, Blase (1988) generated descriptive and conceptual findings on the everyday politics of favoritism in schools. The novel phenomenon of "contrived collegiality" as it relates to relationships among teachers was discussed by Hargreaves (1991). Corbett (1991) traced the impact of parents' actions on the micropolitics of discipline policy at the school level. The "assumptive worlds" of assistant principals and their functions and consequences for site-level politics were analyzed by Marshall and Mitchell (1991). Other studies have examined everyday political relationships between teachers and students (Blase, 1991; Bloome & Willett, 1991; Optolow, 1991); between a department head and teachers (Sparks, 1990); and among a superintendent, principal, and teachers (Kleine-Kracht & Wong, 1991); as well as teachers' political responses to school reform efforts (Noblit, Berry, & Dempsey, 1991).

In parallel fashion, relationships between school principals and teachers have received increased attention in micropolitical research. This emerging stream of research has emphasized the politics of principal control and its adverse consequences for teachers. Anderson (1991) observed that the practice of ideological control by school principals severely limits teacher participation in organizational processes. Blase (1990) investigated the control and protectionist politics of principals and their negative impact on teachers' classroom and schoolwide performance. Ball (1987) linked three "control" styles of school heads (i.e., principals) primarily to negative outcomes such as frustration and fatalism in teachers. Thus far, only Greenfield (1991) has explored the political orientation of open and effective principals in relation to teachers. From observations of both cooperative and consensual processes between one school principal and faculty in an urban elementary school, Greenfield produced brief descriptions of the principal's strategies and their consequences for teachers. He concluded that the principal's power in relation to teachers was derived largely from a common value (i.e., moral) base.

Despite the glimpses of positive and collaborative forms of politics provided by Greenfield (1991), the descriptive and theoretical work on the use of such positive influence strategies by principals and on their impact on teachers is limited. For example, different strategies and techniques were not discussed conceptually in Greenfield's work, nor were teachers' responses differentiated and described in detail. And although Ball (1987) describes three control styles used by school heads, his emphasis is on the negative responses of teachers.

In these and other respects the present study contrasts sharply with related work by Greenfield and Ball. The data presented here (n=836) are a subset of a wide sample of over 1,200 public school teachers; principals' influence strategies and their perceived affective, cognitive, and behavioral impacts on

teachers are specifically described; underlying control and exchange dynamics explaining political interactions are detailed; a new concept of teacher work involvement is presented; and this and other theoretical findings are discussed using elements from Etzioni's (1961, 1975) compliance theory.

Research Methods and Procedures

Data were collected and analyzed from a symbolic interactionist perspective. This perspective recognizes that although structural factors (e.g., organizational, cultural) influence action, the interpretations and meanings that people attach to such factors account for action. In other words, people's capacity for reflexivity has more influence on action than structural factors. The symbolic interactionist perspective views the individual as a social product who is influenced by others but who also maintains distance from others and is able to initiate individual action (Blumer, 1969; Mead, 1934). In contrast to some qualitative research orientations, symbolic interactionism stresses the structure of individual consciousness and perceptions (Blumer, 1969; Tesch, 1988).

Thus the larger study from which this article was drawn employed open-ended questions and focused on the broad question, What are teachers' definitions of the strategies school principals use to influence them? Consistent with inductive-grounded approaches to qualitative inquiry, interpretive data relevant to teachers' perspectives were collected and analyzed to generate descriptive categories and conceptual and theoretical ideas (Bogdan & Biklen, 1992; Bogdan & Taylor, 1975; Glaser, 1978; Glaser & Strauss, 1967).

Allport (1942) suggests that an open-ended questionnaire can be a useful personal document for qualitative research that attempts to focus on the subjective perceptions of individuals. Such an instrument is defined as "any self-revealing document that intentionally or unintentionally yields information regarding the structure, dynamics and functioning of the author's life" (p. xii). A questionnaire is defined as a personal document when the research participants exercise substantial control over the content of their responses. Questionnaires of this type have been employed successfully in other recent research (e.g., Blase, 1986, 1988; Blase & Pajak, 1986; Pajak & Blase, 1989). The research described in this article relies on teachers' perceptions of principals' strategies collected outside of the school settings in which teachers worked. Therefore, the accuracy of the teachers' perceptions cannot be demonstrated here. In fact one would expect the perceptions of people occupying different roles to vary (Blumer, 1969; Bogdan & Taylor, 1975; Mead, 1934).

An open-ended questionnaire, the Inventory of Strategies Used by Principals to Influence Teachers (ISUPIT), was designed to elicit personal meanings regarding the research topic. To develop the first version of the questionnaire, the researcher consulted with a committee of three professors and five teachers. This instrument was piloted with a diverse group of 39 full-time teachers who were graduate students in education at a major state university. Suggestions made by the committee and students were considered in the construction of the final form of the instrument.

The ISUPIT consists of three legal-size pages. On the first page teachers are asked to provide basic background information and to rate their principals with regard to three aspects of leadership—closedness-openness, ineffectiveness-effectiveness, and authoritarian-participatory—on three 7-point scales.

On the two subsequent pages, teachers are asked to provide detailed descriptions of *two* influence strategies used by their school principals. The specific items listed on page 2 and repeated on page 3 are as follows:

> (a) Describe and give a detailed example of a strategy or tactic (overt or covert, formal or informal, positive or negative) that your principal uses frequently to influence what you *do* or *think* in the school or in the classroom.
>
> (b) Describe and give an example of the *effects* (impact) that the strategy has on your thinking and behavior (if any).
>
> (c) Describe and illustrate what you believe to be your principal's *goals/purposes* in using the strategy identified above.
>
> (d) How *effective* is the strategy in getting you to think and do what the principal intended? [7-point scale]. Please explain why.
>
> (e) What *feelings* (if any) do you have about your principal's use of this strategy?

Exploratory-inductive survey research of this nature is designed to produce data-based descriptive categories and theoretical understandings by including a wide variety of research participants (Bogdan & Biklen, 1992; Bogdan & Taylor, 1975; Glaser, 1978; Glaser & Strauss, 1967), thereby maximizing variation in the data (i.e., differences). Because of the open-ended design of the ISUPIT, the substantial time required for completion (about 40 minutes), and the sensitivity of "political" research to practitioners, a mail survey was ruled out. Rather, 14 professors of education administered the ISUPIT between 1989 and 1990 to full-time public school teachers who were taking courses in five on- and off-campus centers located in one southeastern, one northeastern, and one northwestern state. Involvement in this study was voluntary; teachers were instructed not to write their names on the research instrument.

Of the more than 1,200 respondents who completed the ISUPIT, 836 identified their principals as open, effective, and relatively participatory on the 7-point scales provided (means were 5.7, 5.9, and 5.2, respectively). This article presents only that portion of the data, with an emphasis on the affective, cognitive, and behavioral *impacts* of the strategies used by principals. Blase (1993) has examined in detail each of the strategies briefly discussed here. Data related to principals who were viewed as closed and ineffective will be presented elsewhere. Although the decision to focus on findings related to open/effective/participatory principals was based in part on space limitations, a more important consideration was the theoretical and practical significance of this area of inquiry in educational administration.

The primary sample consisted of male (n=172) and female (n=664) teachers from rural (n=292), suburban, (n=443), and urban (n=101) school locations. Elementary (n=335), junior/middle (n=284), and high school teachers (n=217) participated. The average age of teachers was 37; the average number of years of teaching experience was 12. The sample included tenured (n=714) and nontenured (n=122) teachers. Married (n=669), single (n=130), and divorced teachers (n=37) participated. Degrees earned were BA/BS (n=299), MEd/EdS (n=523), and EdD/PhD (n=14).

The teachers described both male (n=497) and female principals (n=339). The mean number of years with the current principal at the time of this study

was 4. The ratio of female principals to male principals is substantially higher than the national ratio. This may in part explain the extensive use of what teachers referred to as "positive" (i.e., acceptable) strategies by principals described in this article. Other research points out that female administrators, in comparison with male administrators, tend to emphasize many of these positive strategies and related goals: support for instruction, visibility in classrooms, cooperation and participation with teachers, and a concern for teacher satisfaction and commitment (Shakeshaft, 1987). However, the present data indicated no differences in use of strategies for effective male and female principals described in the study sample except with respect to two of the "negative" (unacceptable) political strategies: authoritarianism and coercion. A much higher percentage of males was identified with the use of these two strategies.

In general, slight differences were evident when the study sample was compared with the national distribution of teachers in terms of gender, age, marital status, and degree earned. The sample was slightly higher in the percentage of female teachers and slightly lower in the percentage of male teachers as well as average years of experience. The percentage of rural teachers in the sample closely approximated the national distribution, but the percentages of urban and suburban teachers were lower and higher, respectively, in the sample than they are in the national teaching population. Finally, the sample included a lower proportion of elementary teachers and higher proportions of middle/junior and high school teachers than are found nationally (National Education Association, 1983, 1990).

Data from the subsample of 836 teachers were coded according to principles for comparative analysis (Glaser, 1978; Glaser & Strauss, 1967). This inductive procedure consists of comparing each new element encountered in the data with those coded previously in terms of emergent categories, themes, and theoretical ideas. Line-by-line analysis of each open-ended questionnaire page produced a total of 1,323 examples of strategy use. These were grouped into eight major influence strategies, each consisting of several practices (i.e., actions designed to implement strategies).

Comparative analyses of influence strategies generated three types of perceived impacts (i.e., effects) on teachers: affective, cognitive, and behavioral. Each of these three general categories was coded in terms of specific impacts. For example, within the general category of affective impacts were coded the specific impacts of teacher satisfaction, esteem, motivation, and inclusion. The three general impact categories were further analyzed into two larger domains: the instructional-social (student) domain and the noninstructional domain.

Display matrices were constructed for the purpose of synthesizing perceived impacts identified with each influence strategy (Miles & Huberman, 1984). These matrices facilitated numerical analyses of strategies as well as further substantive analyses of impacts. This procedure also permitted the identification of different strategies with similar impacts.

In addition, descriptive matrices were used to identify conceptual and theoretical codes grounded in the data (Miles & Huberman, 1984). For example, these matrices permitted comparisons across strategies and were useful in identifying and refining analyses of emergent ideas such as control, ex-

change, and empowerment. This approach also facilitated comparisons of the study data with the relevant literature, a procedure consistent with grounded qualitative methods (Bogdan & Biklen, 1992; Bogdan & Taylor, 1975; Glaser, 1978; Glaser & Strauss, 1967). Specifically, some of the major findings from the present study are discussed in terms of certain elements from Etzioni's (1961, 1975) compliance theory.

Two questionnaire pages were available for descriptions of strategies. On each of these two pages respondents discussed one influence strategy and frequently several perceived impacts of that strategy. (Some respondents, however, described only one strategy.) Therefore, each completed page was analyzed for only one influence strategy, the strategy identified in item "a" on each questionnaire page. (This point is important, as respondents often alluded to other strategies used by their principal in their descriptions of a particular strategy.) In all, 836 teachers described 1,323 examples of influence strategies used by principals with whom they worked and who they believed were open and effective. These teachers also discussed a total of 1,703 examples of impacts. Throughout this article, *f* refers to the frequency of responses associated with a given influence strategy or impact.

The data related to each influence strategy were also inspected to determine if its use was linked to such characteristics as gender of principal and gender of teacher. No conclusions could be drawn from this analysis.

One researcher analyzed the questionnaire data, a procedure requiring approximately 1,200 hours. In addition, three professors and five doctoral students were consulted when questions arose. As noted, each matrix was designed to display a different segment of raw data, categories of data, and data related to thematic and theoretical codes. As a final check of the researcher's analysis, three doctoral students were trained to examine samples of the study data. This procedure produced an interrater reliability score of .97.

Consistent with the principles for inductive research, all the basic descriptive categories (e.g., strategies and impacts), broader categories/themes (e.g., control, compliance, exchange, empowerment, work involvement), and conceptual ideas (e.g., normative-instrumental leadership, normative instrumental involvement) emerged directly from substantial data appearing on the ISUPIT. No a priori concepts were used to code data. However, several of these emergent ideas (e.g., control, compliance, exchange) are generally similar to ideas already discussed in the extant literature. Given space limitations, brief quotes are used to illustrate selected ideas. (Please note that the above description of research and methods is similar to what appeared in Blase, 1993.)

The Findings

This section begins with a brief overview of the concept of *normative instrumental leadership*, followed by descriptions of the eight major influence strategies and three minor influence strategies identified with this type of leadership. Perceived connections between the strategies and affective, cognitive, and behavioral impacts on teachers are then discussed. The section concludes with a description of *normative-instrumental involvement*, a concept constructed from the study data that integrates the three categories of impacts identified above.

Normative-Instrumental Leadership

The term *normative instrumental leadership*, previously derived inductively from the larger study, captures the everyday influence orientation of open and effective principals toward teachers. Normative-instrumental leadership refers to a political orientation in which control of teachers is a primary goal of principals (although some principals also use an empowerment strategy), and such control is explained in terms of both social exchange and principal-teacher value congruence. In terms of the former, open/effective principals communicate "expectations," overtly and subtly, about both the ends (goals, purposes) and means of teachers' work in schools. Teachers tend to comply with such expectations in return for the explicit and implicit benefits (e.g., rewards, support, goal achievement, opportunities for participation) associated with interactions with principals (Blase, 1993). Exchange was coded in 71% of all data available for the control-oriented strategies to be discussed.

In terms of the latter, the study data suggest that the normative-instrumental orientation of open/effective principals tends to provoke compliance because, on the whole, it is consistent with teachers' professional values and norms in two fundamental respects. First, it is characterized by the use of "normative" influence strategies (i.e., means of influence congruent with teachers' values/norms). Second, it is characterized by the principals' pursuit of "normative" goals (i.e., goals that teachers value and consider legitimate). Thus the term *normative* is used to denote the use of positive (acceptable) influence strategies to achieve goals teachers consider appropriate. For example, it was found that effective principals use predominantly normative influence strategies to achieve normative goals related to innovation in classroom teaching, consideration for students, instructional planning, curriculum development, professional growth, teacher reflection, and collegial collaboration—goals that teachers themselves consider appropriate. Analyses indicated that 84% of all data related to positive influence strategies and goals are normative in the ways described above. Not surprisingly, the three minor negative control strategies (i.e., contrived request for advice, coercion, and authoritarianism) to be described did not reflect value congruence as described above; that is, these influence strategies and the goals identified with them were not considered appropriate by teachers.

The term *instrumental* in normative instrumental leadership is used to emphasize the "pragmatic" link identified in the data between the use of influence strategies by principals and their attempts to elicit teacher compliance to pursue their goals. Put differently, teachers indicated that principals' strategies were specifically designed to influence teachers to achieve principal-determined goals. (See Blase, 1993, for a full discussion of normative-instrumental leadership.)

Strategies

The concept of normative instrumental leadership consists of influence strategies used to achieve the general goal of control in the context of both value congruence and exchange processes vis-à-vis teachers. In fewer cases, empowerment surfaced as a general goal of open/effective principals. Influence strategies identified with the goal of control are described as control-oriented; those identified with the goal of empowerment are considered

empowerment-oriented. Seven major influence strategies and three minor influence strategies were identified with the control orientation (81% of the total data), and one major strategy—involvement in decision making—was identified with the empowerment orientation (19% of the total data).

A strategy was coded as control-oriented when teachers reported that school principals' goals or teachers' means to achieve goals were decided unilaterally and teachers believed they were obligated to comply. Teacher compliance was reported in over 93% of the data related to the control orientation. *Exchange* was coded when teachers reported that strategic interactions with principals were based on transactions of symbolic (e.g., advice, praise) and tangible (e.g., funds, materials) goods; as noted, exchange processes appeared in 71% of available data for all control strategies. Teachers did not associate exchange processes with the empowerment orientation.

A strategy was coded as empowerment-oriented in two ways: *Decision making* (49% of empowerment data) was coded when it was reported principals and teachers jointly assumed authority for determining goals and the means to achieve them, and/or principals empowered teachers individually or collectively to do so. *Authentic request for advice* (51% of empowerment data) was coded when teachers reported they gave opinions on a narrow range of issues defined by principals. Taken together, decision making and authentic request for advice constitute the strategy—involvement in decision making.

The effectiveness rating (i.e., the perceived capacity of each strategy to produce affective, cognitive, and behavioral impacts on teachers) of six of the seven major control-oriented strategies (the exception was formal authority) and the one empowerment strategy were over 6.0 (on rising 7-point scales). The mean effectiveness rating for formal authority was 5.7; for the three minor negative control strategies (i.e., contrived request for advice, coercion, authoritarianism), effectiveness ratings were 3.8, 3.5, and 2.3, respectively.

The Control Orientation

The four positive major influence strategies—rewards, formal authority, support, communication of expectations—and the three minor negative influence strategies—contrived request for advice, coercion, authoritarianism—identified with the control orientation of normative-instrumental leadership are described below.

Rewards (f=300). Two types of rewards, praise (f=268) and material (f=32), were coded in the data. Praise refers to principals' willingness to recognize individuals and entire faculties for their successes, particularly exceptional performance. Terms such as *praises, compliments*, and *gives credit* were used to describe this everyday political strategy. Effective and open principals used formal and informal, public and private, and oral, nonverbal, and written means to praise teachers; praise tended to occur regularly and predictably, most frequently through face-to-face interaction with individuals. Principals also provided teachers with material rewards that carried special significance.

Formal authority (f=161). As a strategy, authority refers to exercising the legitimate rights/powers of the principalship to elicit changes in teachers. *Assigns, states, announces*, and *mandates* were terms teachers used to describe this strategy. The data indicate that although authority was typically exercised unilaterally, it was applied fairly and respectfully. Authority was used widely

to influence teachers' instructional and noninstructional work performance. Compared with all other positive influence strategies, formal authority had lower perceived impacts on teachers.

Support (*f*=137). The data suggest that as a means of everyday influence, principals provide several forms of support to teachers. *Advice and direct intervention* related to instruction was described as one form of support. Advice, both solicited and unsolicited, was given to "help" teachers deal with a wide range of needs and problems.

In addition, teachers reported that to influence them principals offered *administrative support* by attempting to reduce or eliminate factors (e.g., paperwork, faculty meetings) that interfered with their time.

Student-related support refers to a willingness on the part of principals to back up teachers in their decisions regarding student misbehavior and to stand behind teachers in disputes with the parents of students.

Principals also made available *financial and material support*, usually to meet instructional or professional growth goals of teachers. This type of support refers to providing monetary (e.g., tuition for professional conferences) and material assistance to effect changes in teachers' instructional abilities and generally to promote teacher development.

Finally, open and effective principals provided formal (e.g., staff development) and informal (day-to-day, impromptu, casual) *supportive training* to develop practical knowledge and skill in teachers.

Communication of expectations (*f*=112). Although it can be argued that all influence strategies discussed in this article implicitly convey expectations, open and effective principals also spent significant amounts of time clarifying and reinforcing expectations. Terms such as *clarifies*, *explains*, and *informs* were used to describe how principals communicate their expectations.

Visibility (*f*=75). This influence strategy refers to a principal's willingness to spend substantial amounts of time in locations throughout the school—including hallways and classrooms—developing a schoolwide presence, "being available," and taking advantage of opportunities to influence teachers.

Modeling (*f*=115). Modeling and the following influence strategy—suggestion—were seen as slightly less directive than those already discussed. Consequently, teachers allowed themselves slightly more discretion and variation in response to these strategies. Modeling describes actions by principals that dramatize a range of implicit and explicit expectations designed to influence teachers. Three personal characteristics of principals—optimism, consideration, and honesty—were coded as aspects of modeling. Optimism refers to a global and positive orientation; consideration refers to exhibiting a sincere and broad interest in "teachers as human beings"; and honesty refers to a willingness to be straightforward.

Suggestion (*f*=72). This influence strategy, in contrast to more overtly directive strategies, relies heavily on interpersonal diplomacy and facilitative techniques such as questioning. Open and effective principals frequently used the professional literature (e.g., research articles) to effect changes in teachers.

Contrived request for advice (*f*=52), *Coercion* (*f*=28) and *Authoritarianism* (*f*=23). Contrived request for advice describes principals' eliciting input from teachers that would then be ignored as the principals attempted to influence teachers in

Table 1
Strategies Used by Effective Principals to Influence Teachers

Influence strategies
Control orientation:
Rewards (*f*=300)
Formal authority (*f*=161)
Support (*f*=137)
Communication of expectations (*f*=112)
Visibility (*f*=75)
Modeling (*f*=115)
Suggestion (*f*=72)
Contrived request for advice (*f*=52)
Coercion (*f*=28)
Authoritarianism (*f*=23)
Empowerment orientation:
Involvement in decision making (*f*=300)

predetermined directions. Coercion refers to the direct and indirect use of punishment or the threat of punishment to control teachers. Authoritarian practices emphasize preempting teachers' involvement in decisions. It should be emphasized that although these strategies were used by principals considered to be open and effective, these individuals were seen by teachers as relatively less open and less effective than others discussed in the data base. The data also indicate, as previously noted, that these three influence strategies (and the goals identified with them) were not congruent with teachers' professional norms/values and thus they usually failed to influence teachers in positive directions.

The Empowerment Orientation

Involvement in decision making (*f*=300). This strategy constitutes the empowerment dimension of normative-instrumental leadership. To reiterate, this influence strategy consists of two subcategories: decision making and authentic request for advice. To reiterate, decision making describes a situation in which principals and teachers jointly make decisions and/or teachers individually or collectively do so. Authentic request for advice describes a condition in which teachers share their opinions on issues identified by principals.

Principals' willingness to encourage teacher involvement in decision making was discussed by respondents as both a formal and an informal strategy. Formal means included the use of a variety of committee/team structures to elicit teacher participation.

Although teachers' formal involvement in decision making was sometimes direct, more commonly principals limited decision making to key administrative personnel (e.g., department chairpersons). According to the data, teachers were more involved in making decisions when formal participatory structures were available. However, participatory structures notwithstanding, principals continued to influence (although they did not unilaterally deter-

mine) goals, topics, problems, and actions to be taken. This was evident even for teachers who worked in schools they identified as "restructured."

Authentic request for advice was also designed to provide legitimate channels for teachers to express their thoughts and feelings on a range of personal and professional issues. Scheduled and impromptu conversations in which principals solicited input directly from teachers on particular issues were mentioned often by respondents.

The Impact of Principals' Normative-Instrumental Leadership on Affective, Cognitive, and Behavioral Dimensions of Teacher Work Involvement

Three types of impacts—affective, cognitive, and behavioral—(and related sub-impacts) were coded for each of the influence strategies described above. In a later section of this article these three types of work involvement are discussed as one concept—normative-instrumental involvement.

Affective Impacts

Major affective impacts of principals' normative-instrumental leadership include satisfaction, motivation, self-esteem, security, and inclusion. Teachers reported that most influence strategies used by open and effective principals resulted in substantial increases in work *satisfaction*. These included, in descending order of frequency, rewards (100% of responses for this strategy), suggestion (100%), visibility (91%), modeling (86%), support (73%), involvement in decision making (48%), and communication of expectations (35%). Only 22% of responses for formal authority impacted on teacher satisfaction (see Table 2). Terms such as *excellent, happy, glad, comfortable, positive,* and *good* were used to describe feelings of satisfaction. One teacher's remarks about the strategy involvement in decision making (authentic request for advice) illustrate impacts on satisfaction:

> My principal often asks for suggestions on the solution of certain problems and considers each suggestion before making a decision. He is informal and very positive. I feel good about this. I am willing to participate (voluntary) in whatever I can do to enhance the functioning of the school because of his democratic manner of conducting business. I am very willing to give 100% plus.... I feel very positive about working with my principal.

Please note that only high impacts are displayed in Table 2. Given the open-ended nature of the questionnaire used to collect data for this study, if analyses produced a minimum frequency of 35% of impact data available for a particular strategy, this was defined as a high impact.

Teachers also indicated that the use of the positive influence strategies—visibility (100%), suggestion (100%), rewards (95%), modeling (93%), support (63%), communication of expectations (61%), involvement in decision making (55%), formal authority (38%)—enhanced their motivation. *Exalted, inspired,* and *enthusiastic* were words used to denote increases in motivation. For instance, one teacher wrote, "With financial support I could see that I had access to the means of participating in improvement programs. It eliminated the old hang-up of enthusiasm without means of accomplishment." Another noted, "This strategy [visibility] keeps me on my toes mentally because I want to be 'up' and doing my best when my principal is around."

Table 2

*High Impacts of Influence Strategies Used by Effective Principals on Affective Dimension of Teacher Work Involvement

	Affective impacts/Percent of total responses (f) reported for each strategy					
	Satisfaction	*Motivation*	*Self-esteem*	*Security*	*Inclusion*	*Negative Feelings*
Influence strategies						
Control orientation:						
Rewards (*f*=300)	100	95	48		48	
Formal authority (*f*=161)		38				40
Support (*f*=137)	73	63		58(**student)		
Communication of expectations (*f*=112)	35	61				
Visibility (*f*=75)	86	100		43		
Modeling (*f*=115)	91	93				
Suggestion (*f*=72)	100	100				
Contrived request for advice (*f*=52)						93
Coercion (*f*=28)						
Authoritarianism (*f*=23)						
Empowerment orientation:						
Involvement in decision making (*f*=300)	48	55	45	43		

Note. *When a minimum of 35% of the available data for a strategy are identified with an impact, this is defined as high.
**=type of support.

In addition, the use of rewards (48%) and involvement in decision making (45%) were strongly identified with perceived impacts on self-esteem. Terms that appeared frequently in the data in relation to esteem were *confident, loved, valued, respected, recognized,* and *appreciated.* The comments of one teacher are typical:

> My principal gives a good amount of praise to the teachers at my school—whether it is a note to us, an announcement of our accomplishments to all other teachers, or an individual conference in which she thanks us for the wonderful job we're doing ... My principal's actions have had a very positive effect on my self-esteem and confidence ... I try to do what he expects. If a person is good to me I will always be good to him ... I am more apt to volunteer for her projects that she needs help with at school than if I hadn't been praised.

According to the respondents, two influence strategies—support (student-related) (58%) and visibility (43%)—provoked feelings of security. Words used to describe security included "at ease, not threatened," "relieved," "secure," "free." A teacher noted, "His support is invaluable.... I am relieved knowing he will be there for me." Feelings of inclusion were discussed for the use of rewards (48%) and involvement in decision making (43%). The comments of two teachers are illustrative: "I see myself as an important member of the team." "I am more a part of the school ... and [have] a determination to make things work."

In a few instances, the use of formal authority (9%) and communication of expectations (7%) were viewed as generating mixed feelings in teachers. One teacher, for example, discussed her response to the use of formal authority: "At first I was very upset but then I came to realize that if just one student learns it will be worth it.... [However,] it would have been better if teachers had been involved in the preplanning." Mixed feelings resulting from communication of expectations were described by another: "I feel stress and overwhelmed yet I am glad that those teachers who are not performing are now forced to put up or shut up."

Finally, only the influence strategy of formal authority was frequently identified with negative feelings. Here 40% of the feeling data related to this strategy were decidedly negative. Teachers' negative reactions to formal authority were associated with its perceived overtly controlling, quasi-coercive, and unethical nature; consequently, the strategy was frequently seen as incongruent with teachers' professional norms/values. Terms such as *resentful, angry, bullied, intimidated, tense, guilty, depressed, frustrated,* and *unpleasant* were used to describe feelings. The remarks of one teacher describe the adverse consequences that may occur for those who viewed principals' use of formal authority negatively:

> Last year the principal required us to have silent lunch in the lunchroom three days a week. No one was allowed to talk ... the whole time was unpleasant. Our principal started this because it was too loud in the lunchroom and we were told that if behavior improved ... we would gradually eliminate some of the days ... No days were ever eliminated ... I felt she [the principal] did us a disservice by never backing off. If the topic of taking away quiet days was mentioned she would consider our suggestions reluctantly but no changes were ever made. She

always mentioned that some people weren't enforcing the rule and we all must "pull together."

The three negative strategies of contrived request for advice, coercion, and authoritarianism (respectively) produced only negative feeling states in teachers.

Cognitive Impacts

The major cognitive impacts of principals' leadership are awareness and reflection. The influence strategies (in descending order of frequency) of visibility (100% of responses for this strategy), suggestion (100%), support (training) (73%), communication of expectations (57%), and modeling (54%) were strongly associated with increases in teacher awareness (see Table 2). Formal authority was linked to awareness in 24% of the responses for this influence strategy. Most frequently awareness was described in relation to becoming cognizant of the academic and social needs and problems of students ("I have become aware of the specific needs of our students ... love, discipline, caring, self-esteem and physical needs."). To a lesser extent, teachers reported greater awareness of school-wide issues. (Only high impacts are presented in Table 3.)

Most positive influence strategies discussed in this article were seen as having an impact on teacher reflection to a greater or lesser degree. Reflection refers to the development of an improvement- and problem-centered orientation primarily toward the classroom. For example, teachers thought more strategically about teaching techniques in dealing with students' academic and social needs: "I am constantly trying to think of strategies that emulate my

Table 3
*High Impacts of Influence Strategies Used by Effective Principals on Cognitive Dimension of Teacher Work Involvement

	Cognitive impacts/Percent of total responses (f) reported for each strategy	
	Awareness	*Reflection*
Influence strategies		
Control orientation:		
Rewards (*f*=300)		
Formal authority (*f*=161)		
Support (*f*=137)	73(**training)	
Communication of expectations (*f*=112)	57	
Visibility (*f*=75)	100	51
Modeling (*f*=115)	54	
Suggestion (*f*=72)	100	70
Contrived request for advice (*f*=52)		
Coercion (*f*=28)		
Authoritarianism (*f*=23)		
Empowerment orientation:		
Involvement in decision making (*f*=300)		76

Note. *When a minimum of 35% of the available data for a strategy are identified with an impact, this is defined as high.

principal's actions." A teacher described the impact on instruction of one principal's use of suggestion vis-à-vis the professional literature:

> Our principal uses handouts of research and pertinent information on subjects being taught or discussed at school. He also places in our mailboxes little thought provoking ideas concerning teaching. Coupled with the known knowledge of his thinking ... this lets me know what his expectations are and that I should show signs of professional growth and improvement ... It causes me to analyze my own teaching and evaluate whether I can use the ideas.... I try to include new teaching strategies and/or information in my planning for instruction and my actual instruction.

Some teachers also reported becoming more introspective about personal problems that interfered with their teaching. Less often, increases in teacher reflection were described for routine concerns such as the effective use of available resources and instructional equipment.

However, further analysis of cognitive impacts data pointed to startling differences among positive influence strategies. In fact only three strategies—involvement in decision making (76%), suggestion (70%), and visibility (51%)—appeared to have high impacts on teacher reflection. Other positive strategies were cited by fewer teachers as having an impact on reflection, including communication of expectations (33%), support (18%), modeling (16%), formal authority (13%), and rewards (6%). For the three high impact strategies (i.e., involvement in decision making, suggestion, and visibility), the data suggest that there are several contributing factors: principal trust/respect (for teachers); formal and informal opportunities for expression, coupled with low risk (e.g., of ridicule, criticism) and acceptance; increased levels of teacher meaningful responsibility and authority; sense of efficacy; and collegial interaction. The three negative minor influence strategies discussed in this article had no perceived positive impact on teacher reflection.

Other cognitive outcomes noted less frequently in the data were identified primarily with particular strategies. Use of rewards affected school culture (8%) ("The atmosphere of the school is ... full of expectations. We expect success and achievement and we find these goals to be achieved.") and teachers' desire for participation (12%) ("I desire to participate in school events and support all aspects of the school."). Communication of expectations (6%) was associated with conformity in thought ("I feel that it is expected of me to closely align my thoughts and choices to that of my principal.") and visibility (16%) contributed to teachers' sense of responsibility ("I am aware that I am accountable for my actions."). Involvement in decisions (19%) promoted trust in the system ("I feel there is a way to channel ideas and concerns ... there is a forum for my frustrations or trouble areas that arise."). Support (advice/intervention) (17%) was linked to increases in respect for principals ("I know that he cares ... I see his wisdom in the words and attitudes he expresses.").

Finally, formal authority was seldom identified with positive cognitive responses. Even when teachers responded positively on a behavioral level to the use of formal authority, their cognitive responses were often decidedly negative. As one teacher observed, "Unusual [frequent] criticism makes me less conscientious about living up to my responsibility." Contrived request for

advice, coercion, and authoritarianism were not related to positive cognitive impacts.

Behavioral Impacts

One general theme—compliance (i.e., teachers responded in ways intended by principals)—was derived from substantial behavioral impact data. The link between the use of positive influence strategies by school principals and behavioral compliance in teachers was often discussed explicitly: "I make every effort possible to meet his expectations ... I try to do my best ... live up to his comment ... please the principal." According to the data, teachers traded their compliance for the implicit and explicit rewards (benefits) associated with the principal's use of strategies and, in much fewer cases, for the avoidance of punishment or sanctions. Compliance was evident in 95% of the data coded for five of the major influence strategies. (Only high impacts are presented in Table 4.)

Interestingly, teacher compliance was reported almost as frequently (93% of data) as a response to principals' use of the less directive major strategies of modeling and suggestion. A teacher reported:

> My principal leads through example and shows that hard work and fun in learning go hand in hand.... Her goals are to build the self-esteem and self-confidence of students in a supportive way. She frequently visits my class while we are doing laboratory activities and participates with the students by using the equipment, asking questions and asking me questions.... I try to conform to what I think she wants. I start thinking more positively about my class and my role as a teacher. I respond by being more open with my students and I try to develop the same rapport with them that she has.

In addition, the data strongly suggest that the unwillingness on the part of teachers to exercise greater decisional discretion in both the classroom and the school is related directly to their assumption that principals "expect" compliance and a relatively submissive attitude toward administrative authority. As one teacher explained, "He [the principal] is always the boss ... You don't forget this." However, the use of contrived request for advice, coercion, and authoritarianism failed to provoke significant levels of compliance in teachers. Instead, avoidance and passive forms of resistance were common. Compliance was evident in only 26% of the data for these strategies. The defensive tone of these data is revealing: "I find myself trying to be sure that he sees me before 7:45 a.m. and after 3:45 p.m. to 'prove' that I put in the correct hours each day"; "I felt obligated to do some things ... but it is out of fear of retaliation not a sense of good will."

A second broad theme, referred to here as work involvement, was gleaned from the behavioral impact data. This theme includes both a quantitative and qualitative dimension (i.e., degree of involvement as well as type of involvement). Regarding the former, a teacher commented, "I work harder. This means I don't just put in eight hours a day. I work until I am finished." For the latter, a teacher notes, "I try to do a better job ... work harder to make a unit more successful."

The data indicate that teacher involvement in work varies in relation to principal influence. Three behavioral subcategories—student and classroom

Table 4
High Impacts of Influence Strategies Used by Effective Principals on Behavioral Dimension of Teacher Work Involvement

	Behavioral impacts/Percent of total responses (f) reported for each strategy											
	Student/Classroom							*Professional impacts*				*Negative impacts*
	Consideration for students	*Innovation and creativity*	*Instructional planning*	*Instructional time on task*	*Monitoring student learning*	*Curriculum follow through*	*Student discipline*	*Relationship with principal*	*Professional growth*	*Relationship with parents*	*Collegial collaboration: Expression/Compliance with/ procedures/Decision making/Follow through/ Compromise/Support for decision*	*Avoidance*
Influence strategies												
Control orientation:												
Rewards (f=300)	52	38	37					93	47			
Formal authority (f=161)										37		
Support (f=137)	(**student) 63	(**financial) 40					(**student) 100	84	73			
Communication of expectations (f=112)	46			39			50	41	36			
Visibility (f=75)	42		42	53	39	35		56	39			
Modeling (f=115)	41							89	72			
Suggestion (f=72)	37	65						91	83			89
Contrived request for advice (f=52)												
Coercion (f=28)												
Authoritarianism (f=23)												
Empowerment orientation:												
Involvement in decision making (f=300)								57	87			40

Note. *When a minimum of 35% of the available data for a strategy are identified with an impact, this is defined as high.
**=type of support.

impacts, professional impacts, and negative impacts—were coded for work involvement. Strong student and classroom impacts were observed in the data. Six influence strategies—support (student) (63% of responses for this strategy), rewards (52%), communication of expectations (46%), visibility (42%), modeling (41%), and suggestion (37%)—were linked to increases in consideration for students. This perceived impact was primarily discussed in the context of addressing social-emotional needs; for example, teachers reported being "more open and patient," exhibiting greater concern for students' self-esteem, and "including positive statements in ... explanations of grading policies." One teacher wrote,

> When discussing the needs of a student with disciplinary problems the principal asks us what we think can be done to prevent the student from getting into difficulty.... His approach is non-threatening. The principal will gently suggest ways to keep that student in class or in school. He points out the progress the student has made, the ability level of the student, and what the student probably encounters at home ... His approach makes me more aware and more sensitive to the student's needs and helps me realize ways I can better serve the student's needs. I change my approach to a more sympathetic one to help the student.

The influence strategies of suggestion (65%), support (financial-material) (40%), and rewards (38%) were perceived to provoke noteworthy increases in classroom innovation and creativity. The data imply that the use of these influence strategies provided encouragement to teachers and reduced some of the risks they commonly associate with instruction: "I can try and experiment with new strategies introduced at conferences without fear of risking my 'status' as a good teacher"; "I will give things a try because I do not feel threatened if it fails."

Respondents indicated that various influence strategies were viewed as having other substantial behavioral impacts on students and the classroom. Visibility (42%) and rewards (37%) related to increases in instructional planning; visibility (53%) and communication of expectations (39%) were identified with increases in instructional time on task in classrooms. The strategy of visibility strongly increased monitoring of student learning (39%) and curriculum follow-through (35%). Support (student-related) (100%) and communication of expectations (50%) were viewed as strengthening teachers' abilities to deal with student discipline problems:

> I am motivated to do a good job for my principal because he is always willing to help me with any problems I have and supports me in my decisions. He helped me deal with some students who were disruptive and rebellious ... I am a first year teacher and I'm learning how to deal with discipline problems. He let me know that he trusts my judgment and does not mind helping ... I feel very confident and relaxed about making decisions in the classroom. He wants me to be an effective teacher who has control of my class.... I like what he does.

In a few instances, teachers disclosed that support (advice-intervention) (8%) promoted the development of teachers' crisis intervention skills; and visibility (7%) was connected to increases in teacher authoritarianism ("I tend to be more authoritarian in the classroom because at any time a visit may be paid and I like to make sure that I'm always in control of myself and my

students."). Only one strategy—suggestion (11%)—was seen as expanding slightly teachers' discretion in the classroom: "I weigh the suggestion and if I feel a change is warranted I do that. If there is an area where I feel I am more knowledgeable about the content than he is—then I do it my way."

Most principals' influence strategies were also seen as having a significant impact on professional aspects of teachers' behavior: specifically, relationships with principals and teachers' professional growth. The following influence strategies had high respective impacts: rewards (93%, 47%), suggestion (91%, 83%), modeling (89%, 72%), support (84%, 73%), involvement in decision making (57%, 87%), visibility (56%, 39%), and communication of expectations (41%, 36%). Formal authority (21%, 15%) had lower respective perceived impacts on relationships with principals and professional growth. Formal authority (37%) also was viewed as enhancing relationships with parents. Less frequently teachers reported that the use of rewards (17%) affected teacher volunteerism ("I am more apt to volunteer for projects she needs help on."). Modeling influenced teacher compliance with bureaucratic requirements including punctuality (6%) (i.e., arriving at work on time) and professional dress (9%). Formal authority also related to improvements in teacher punctuality as well as attendance at meetings, compliance with rules, and observance of deadlines. About 15% of the responses for formal authority were reported for each of these impacts.

Involvement in decision making was associated with several high impacts not described for other strategies. All reported impacts resulted from involvement with school-level decision making. No impacts on the classroom were discussed for this strategy. Each of the following impacts was reported roughly in 40% of the data for this strategy. One teacher described impacts on expression:

> Our principal organized a Leadership Management Team (L.M.T.) in our school. This team is composed of representatives from all grade levels.... The team discusses concerns and policies for the school. Each representative presents these from their grade level and returns to the grade level to discuss results.... I think I am being treated like a valuable, intelligent professional. I participate more and feel free to express my opinion ... even on controversial matters.

Other high impacts identified with involvement in decision making were compliance with procedures for decision making ("Now I follow the process of change when I have a complaint or concern."), and follow-through ("I know others depend on me to bring back information."), compromise ("I am more willing to compromise when we don't agree; he [the principal] can see my viewpoint and I can see his.") and support for decisions ("I am supportive of ideas that evolve through dialogue and will implement them whether I have been directly involved in the development of the task or not.").

Not surprisingly, teachers' data indicated that contrived request for advice, coercion, and authoritarianism were linked to only negative behavioral responses, predominantly avoidance of principals (89%) ("I attempted to avoid interaction with my principal ... keep a low profile."). A few instances of passive resistance (but no instances of overt resistance) to principals appeared in the data.

Summary and Additional Analyses of Impacts

All positive influence strategies resulted in predominantly positive affective impacts, especially increases in teacher satisfaction. Of the positive influence strategies employed by effective principals, teachers disclosed that only formal authority had minimal positive impacts on satisfaction; this influence strategy was not linked to other positive affective impacts. Various combinations of strategies were related to increases in motivation, esteem, feelings of security, and inclusion.

The study respondents identified three positive strategies with strong cognitive impacts, particularly increases in reflection. However, formal authority and rewards were not highly related to increases in teacher reflection. It appears that formal authority was used to elicit compliance to routine operational expectations (e.g., meeting deadlines, rule compliance) and to expectations regarding professional dress. Apparently, neither of these forms of compliance requires ongoing reflective thinking by teachers. Even when formal authority was discussed in relation to achieving important academic and social goals related to students, it was often viewed as aggressive and blatantly controlling. Consequently, teachers indicated that they did not identify with and internalize the goals and values of principals who relied heavily on this strategy. Stated differently, teacher compliance to formal authority is often without strong positive affective and cognitive corollaries—it is primarily a surface-level behavioral response on the part of teachers. As such, formal authority, at least in the teachers' perspective, fails to produce identification with and internalization of principals' goals/values and teacher work involvement that reflects deeper levels of commitment.

Principals' use of positive influence strategies directly affected behavioral compliance; that is, teachers responded to principals' use of strategies and expectations (goals) in the desired ways. Explicitly and implicitly the data point out that the use of positive influence strategies (except formal authority) was linked to major impacts on relationships with principals and also with teachers' professional growth. Increases in time devoted to work and, more concretely, greater consideration for students are only two ways in which teachers frequently complied with principals' expectations. In addition, teachers reported idiosyncratic behavioral responses for principals' use of such strategies as rewards, formal authority, modeling, visibility, and involvement in decision making.

The data also indicate that teachers rarely complied with the expectations of principals who used contrived requests for advice, coercion, or authoritarianism. When cognitive and behavioral compliance did occur, the study data indicate that behavioral compliance often exceeded cognitive agreement: "I will do what is asked even if I question his intentions."

Instructional-Social Versus Noninstructional Domains

Continued analyses indicated that the three types of perceived impacts—affective, cognitive, behavioral—could be categorized in terms of an instructional-social (student-related) domain or a noninstructional domain. Teachers discussed significantly more data for the former domain. This analysis indicates that major affective impacts such as satisfaction, motivation, esteem, and security are usually associated with the instructional-social domain as are

major cognitive impacts including awareness and reflection. Behavioral impacts such as consideration for students, creativity/ innovation, time on task, instructional planning, monitoring of student outcomes, curriculum follow-through, and student discipline were also discussed predominantly as instructional-social domain issues. In total, the data identified with this domain account for about 69% of all frequency impact data.

In contrast, teachers' identified relatively few data (31% of frequency impact data) with the noninstructional domain. The only affective impact is inclusion: cognitive impacts associated with this domain include desire for participation, trust in the system, and respect for principals. Noninstructional behavioral impacts were observed for volunteerism and relationships with principals and parents. Use of formal authority was discussed for some idiosyncratic noninstructional behavioral impacts (e.g., punctuality, attendance at meetings); idiosyncratic impacts associated with involvement in decision making were compliance with decision procedures, follow-through, compromise, and support for decisions. The influence strategies of contrived request for advice and authoritarianism were tied to reductions in faculty involvement in decision making.

Indeed, the study data argue that the influence strategies described in this article tend to result in positive impacts on teachers and, overall, such impacts represent various forms of teacher compliance with principals and their goals. As the above analysis suggests, this was especially apparent for the instructional-social domain, which focused on teacher compliance directly in terms of working with students. More abstractly, it has been argued that teacher compliance occurred largely because of a perceived congruence between principals' strategies and goals, and teachers' professional norms/values. This conclusion was especially salient in data related to the instructional-social domain; however, as indicated earlier, it was also reflected in 84% of the total data related to the use of positive strategies. One teacher's comments illustrate the importance of congruence between principals' use of positive influence strategies and teacher compliance.

> A strategy that my principal frequently uses to influence what I do is one that is easily identified. She comes to me frequently, touches me on the shoulder and verbally expresses how thankful she is to have me as part of her staff. She makes me feel needed and important to the school.... I perform with confidence and boldness and seek to please her.... All people like to be complimented. Praise is often all we need to motivate us to be the best that we can be ... I agree with this strategy. I treat my students the same way.

In contrast, the data argue that teacher compliance seldom occurs when there is lack of perceived congruence between principal strategies/goals and teachers' professional norms/values. Specifically, data available for the strategies of coercion, authoritarianism, contrived request for advice, and to some extent for formal authority suggest that when principals fail to employ appropriate influence strategies and/or pursue inappropriate goals, teacher compliance is problematic. As suggested earlier, avoidance was the teachers' major response to several of the aforementioned influence strategies.

Normative-Instrumental Involvement: An Emergent Concept

The three types of impacts described above—affective, cognitive, behavioral—represent different ways teachers involved themselves in work, quantitatively and qualitatively, in response to principals. This point is supported by teachers' descriptions of impacts related directly to achieving principal-determined work-related goals (i.e., broadly, the instructional and social goals discussed above). In most instances, teachers responded to principals' use of positive influence strategies/goals by increasing their involvement in work. Therefore, in accord with guidelines for grounded theory research, the concept of *normative-instrumental involvement* was constructed to reflect these three dimensions of teacher work involvement (Glaser, 1978; Glaser & Strauss, 1967).

Although normative-instrumental involvement is primarily a form of teacher compliance, it is considered normative because it is grounded in the congruence between the moral-value dispositions of teachers and those reflected in the influence strategies and goals of principals. This congruence was especially evident in data describing principals' instructional and social goals for students. Teachers' willingness to comply with principals by increasing their involvement in work coincided with their inherent interest in achieving principal-determined goals. Stated differently, principals' ability to influence teachers is largely a function of teachers' willingness to be influenced in certain ways and toward certain ends. This congruence between teachers and principals was coded in 84% of the frequency data for the positive influence strategies.

The term *instrumental* as used in the concept of normative-instrumental involvement underscores the focus of this form of involvement: It is designed to achieve work-related goals, particularly those identified with the social/instructional domain of work with students.

It is argued above that normative-instrumental involvement is contingent on exchange processes initiated and sustained by school principals over time. In part such processes, teachers report, contribute to the development of particular affective, cognitive, and behavioral dispositions. The use of positive influence strategies by principals is linked to specific explicit and implicit benefits such as teacher satisfaction, esteem, praise, support, and goal achievement. In return for these and other benefits (most of which are symbolic) teachers tend to comply with principals' expectations by expanding their cognitive (e.g., reflection) and behavioral (e.g., time for instructional planning) involvement in work. As noted earlier, exchange processes appeared in 71% of the frequency data. The following quotation illustrates the prominence of symbolic exchanges (e.g., praise for compliance) in the study data:

> Mr. ____ gives me positive reinforcement through written feedback for things he felt were worthy of praise—regular parent—teacher conferences, helping other teachers with problems, demonstrating a positive attitude toward students both in and out of the classroom ... I am encouraged to work to the greatest of my potential because I want to do my best and please the principal. I would feel guilty if I did not do extra work and go the extra mile for my principal.

The importance of exchange process to social interaction in general (e.g., Blau, 1964; Homans, 1958; Thibault & Kelley, 1959) and micropolitical interaction in particular (e.g., Bacharach & Lawler, 1980; Ball, 1987; Hoyle, 1986) is well established in the literature. Exchange has also been discussed as a critical

component of understanding relationships between leaders and followers (e.g., Burns, 1978; Hollander, 1978; Jacobs, 1970). Although comparisons failed to yield a consistent fit between the present data and any specific approach to social exchange, certain basic ideas associated with exchange theories in general were evident in the data. Typically, exchange theories assume that interaction between individuals and groups in social settings consists in large part of transactions of tangible and intangible (e.g., acceptance, conference) valued goods. Individuals and groups engage, explicitly and implicitly, in a process of mutual accommodation to secure perceived rewards and benefits. Exchange processes are viewed as producing both social debt and a sense of obligation in others to reciprocate. Among other things debt creation and reciprocation promote social exchanges between and among individuals and groups over time (Blau, 1964; Gouldner, 1960; Homans, 1958; Thibault & Kelley, 1959). It is these general ideas about social exchange that emerged in the present findings.

In sum, it should be emphasized that normative-instrumental involvement, although fundamentally a form of compliance, is experienced quite positively by teachers. On one hand, it is a willing rather than a coerced act and based on a decision to accept as proper a given expectation (Pratt, 1988). On the other hand, normative involvement is contingent on exchanges (particularly "symbolic" exchanges) initiated and maintained by principals. In general, principals articulate *their* visions, set *their* goals, explain *their* expectations, communicate *their* values, and in large part determine the means to achieve such ends. Teachers are encouraged to "buy into the principal's agenda" in exchange for explicit and implicit rewards associated with the use of positive influence strategies. For these reasons, the exercise of normative-instrumental leadership seems to expand what Barnard (1938) referred to as administrators' "moral" authority and teachers' "zone of indifference," that is, the degree to which the expectations (goals/purposes) of school principals are considered legitimate and thus acceptable and teacher compliance is voluntary.

Although normative-instrumental leadership promotes internalization of principals' goals and related values, teachers seldom indicated that they achieved high levels of empowerment and developed leadership qualities as a result of working with effective principals. To reiterate, decision making was coded in the data when it was evident that teachers were involved in actually making decisions collaboratively or when teachers were individually or collectively authorized to make decisions. Only one influence strategy—involvement in decision making—was linked to the development of teacher empowerment. However, only 51% of the data for this strategy were consistent with the concept of decision making described here. Teachers' advisory role in decision making was underscored in 49% of the data for this strategy. In other words, the data suggest that teachers are frequently limited to giving opinions on a narrow range of issues defined by principals, whereas principals retain decisional authority and responsibility. On the whole, it appears that normative-instrumental involvement is "good subordination" and followership rather than the type of professionalism and empowerment to which many prominent educators aspire (e.g., Schlechty, 1990). Plainly, normative-instrumental involvement is consistent with the hierarchical structure of school organization.

Discussion: Theoretical Implications

Findings from the present study argue that the influence strategies of principals have a positive impact on teachers' work involvement. Moreover, such involvement is founded on a strong moral/value commitment to work on the part of teachers. In accordance with guidelines for grounded theory inquiry, the relevant professional literature was reviewed to determine specific linkages between this central finding and the theoretical literature focusing on power/influence, particularly relationships between superordinates and subordinates in organizations (Glaser, 1978; Glaser & Strauss, 1969). Several theorists have emphasized the importance of value congruence between subordinates and superordinates (or leaders and followers) (e.g., Burns, 1978; Barnard, 1938; Henderson, 1981; Mitchell & Spady, 1983). However, because of direct similarities with the present data, Etzioni's (1961, 1975) compliance theory was selected for purposes of brief comparison. This theory, which focuses on relationships between superordinate power and subordinate involvement, has been used primarily to classify organizations. However, it contains certain ideas directly relevant to the central theoretical finding that emerged from the present study.

An initial comparison was made between the influence strategies discussed in this article and the concepts from Etzioni's power model, an important component of compliance theory. (In part, the French and Raven's (1968) model was not used for comparative purposes here because of Bacharach and Lawler's [1980] criticism that this model is conceptually inconsistent and mixes bases of power and sources of power.) This model consists of four means of control: coercive (use or threat of physical sanctions); remunerative (material resources, salaries, services, commodities); normative (symbolic: manipulation of prestige, rituals, symbols, love, acceptance); and knowledge (control of information). (In Etzioni's earlier model knowledge was conceptualized as normative control; Bacharach & Lawler, 1980, expanded the model.) This analysis indicates that although open and effective principals use a range of influence strategies to influence teachers, such strategies rely most heavily on positive "normative," that is, symbolic forms of power. Only the strategies of coercion, authoritarianism, and in part formal authority were defined as coercive and overtly controlling by teachers. According to the data, teachers seemed to experience contrived request for advice as a negative form of normative power. Slight evidence of remunerative power was found in data related to the strategy of rewards. For a full discussion of this analysis, see Blase (1993).

The relationship between compliance theory (Etzioni, 1961, 1975) and the present study was especially strong with regard to principals' use of power and teacher work involvement. Etzioni theorizes that specific types of superordinate power are congruent (compatible, effective) with certain forms of subordinate involvement. He hypothesized that superordinates' use of coercive power is consistent with alienative involvement (i.e., negative, hostile); remunerative power is congruent with calculative involvement (i.e., neutral feelings, interest in material benefits); and normative (i.e., symbolic) power is congruent with moral/value involvement (i.e., based on identification with authority, internalization of norms). Here congruence refers to compatibility of superordinates' use of power and subordinates' involvement. Each requires

the other. To illustrate, the effective application of normative (i.e., symbolic) power requires moral/value-based involvement (i.e., commitment) by subordinates; variations in control can be expected to provoke variations in moral involvement, and vice versa.

The present study data are generally consistent with Etzioni's (1961, 1975) hypothesis regarding the use of normative (i.e., symbolic) power by superordinates and moral involvement on the part of subordinates. Analyses of affective, cognitive, as well as behavioral dimensions of work involvement discussed throughout this article provide compelling evidence that principals' use of normatively based strategies provokes deeper and more intense levels of moral-based involvement by teachers. At the same time, principals' use of normative-symbolic power requires, in varying degrees, moral involvement on the part of teachers. Etzioni (1975) further contends that this type of involvement is often critical to the realization of symbolic goals in normative organizations.

Indeed, the importance of teachers' strong moral involvement to achieving educational and social goals with students is well established in the empirical literature (Dreeben, 1968; Lortie, 1975; Rosenholtz, 1989; Waller, 1932; Wynne, 1987). However, to date no studies have been published that are consistent with Etzioni's general hypothesis as it relates to school principals' use of normative power and teachers' moral involvement. In education, relevant studies have concentrated on teacher-student relationships. Only Greenfield's (1991) study of one school principal is suggestive along these lines. Also, empirical research on how principals influence teachers indicates that the use of normative forms of power may result in positive teacher outcomes, for example, collaborative relationships, loyalty, satisfaction, professional conduct, and commitment (Blase, 1987; Hanson, 1976; Hoy & Brown, 1986; Johnston & Venable, 1986; Leithwood & Jantzi, 1990; Tartar, Hoy, & Bliss, 1989).

A Final Note

This article discusses the positive influence strategies used by open and effective principals and their impact on teachers. Such principals employ normatively based strategies to achieve normative goals. Effective principals are predominantly control-oriented: They directly and indirectly define both the goals and the means to achieve goals for teachers. And although such a control orientation results in increased levels of involvement by teachers, this involvement is largely a form of compliance. According to the present study, even effective principals rarely promote significant levels of teacher involvement in school decision making.

It is also important to note that only three strategies discussed in this article—visibility, suggestion, and involvement in decision making—resulted in a significant increase in teacher reflection. The strategies of support, modeling, and communication of expectations had only small impacts on reflection. Specifically, the former three (high impact) strategies were linked to increases in teacher authority, efficacy, collegial interaction, and openness; such strategies provoked complex, creative, and strategic thinking on the part of teachers. It is precisely such forms of thinking that enhance teachers' capacity for leadership in restructured schools (Conley, 1991) and improve their chances

of dealing successfully with students' academic and social needs (Sparks-Langer & Colton, 1991).

In discussing the current educational reform movement, Conley (1991) notes that a compliant orientation in teachers vis-à-vis school administrators must be replaced with greater teacher authority and responsibility in school-wide strategic and operational decisions. Although the research of Conley and others has provided powerful rationales for viable forms of teacher participation, this has typically not occurred in most American schools. Generally, teachers experience considerable "decision deprivation," particularly in the organizational domain of schools (Bacharach, Bamberger, Conley, & Bauer, 1990). In fact, recent empirical studies of restructured schools have produced startling evidence of this problem (Imber & Duke, 1984; Malen & Ogawa, 1988).

During the last several years there has been a consistent call in restructured schools for the kind of leadership discussed by Burns (1978). Such leadership will require visionaries: women and men who will transform educational institutions through authentic forms of teacher involvement in decision making (Schlechty, 1990). To support efforts to restructure schools, university programs in educational administration, particularly those dominated by a control ideology, will be required to reexamine their perspective on school leadership. Similarly, changes in the perspectives of teacher education programs that explicitly and implicitly foster a subordinate and compliant orientation in teachers should be examined. In both cases, consideration of critical perspectives that emphasize an understanding of the relationship between dominant societal structures, hierarchical/bureaucratic ways of managing, and the inequitable distribution of power, knowledge, and resources in schools can be valuable (see, e.g., Apple, 1982; Bates, 1982; Foster, 1986; Giroux, 1981; Yeakey, 1987).

References

Allport, G. (1942). *The use of personal documents in psychological science.* New York: Social Science Research Council.

Anderson, G. (1991). Cognitive politics of principals and teachers: Ideological control in an elementary school. In J. Blase (Ed.), *The politics of life in schools: Power, conflict, and cooperation* (pp. 120-138). Newbury Park, CA: Sage.

Apple, M. (1982). *Education and power.* Boston: Routledge & Kegan Paul.

Bacharach, S.B., Bamberger, P., Conley, S., & Bauer, S. (1990). The dimensionality of decision participation in educational organizations: The value of a multidomain evaluative approach. *Educational Administration Quarterly, 26,* 126-167.

Bacharach, S.B., & Lawler, E.L. (1980). *Power and politics in organizations: The social psychology of conflict, coalitions, and bargaining.* San Francisco: Jossey-Bass.

Ball, S.J. (1987). *The micro-politics of the school: Towards a theory of school organization.* London: Methuen.

Barnard, C.I. (1938). *The functions of the executive.* Cambridge, MA: Harvard University Press.

Bates, R.J. (1982, March). *Towards a critical practice of educational administration.* Paper presented to the annual meeting of the American Educational Research Association, New York.

Blase, J. (1986). A qualitative analysis of sources of teacher stress: Consequences for performance. *American Educational Research Journal, 23,* 13-40.

Blase, J. (1987). Dimensions of effective school leadership: The teachers' perspective. *American Educational Research Journal, 24,* 598-610.

Blase, J. (1988). The politics of favoritism: A qualitative analysis of the teachers' perspective. *Educational Administration Quarterly, 24,* 152-177.

Blase, J. (1990). Some negative effects of principals' control-oriented and protective political behavior. *American Educational Research Journal, 27,* 727-753.

Blase, J. (Ed.). (1991). *The politics of life in schools: Power, conflict, and cooperation*. Newbury Park, CA: Sage.

Blase, J. (1993). The micropolitics of effective school-based leadership: Teachers' perspectives. *Educational Administration Quarterly, 29*, 142-163.

Blase, J., & Pajak, E. (1986). The impact of teachers' work-life on personal life: A qualitative analysis. *Alberta Journal of Educational Research, 32*, 307-322.

Blau, P.M. (1964). *Exchange and power in social life*. New York: Wiley.

Bloome, D., & Willett, J. (1991). Toward a micropolitics of classroom interaction. In J. Blase (Ed.), *The politics of life in schools: Power, conflict and cooperation* (pp. 207-236). Newbury Park, CA: Sage.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice-Hall.

Bogdan, R., & Biklen, S. (1992). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.

Bogdan, R., & Taylor, S. (1975). *Introduction to qualitative research methods: A phenomenological approach to the social sciences*. New York: Wiley.

Burns, J. (1978). *Leadership*. New York: Harper & Row.

Conley, S. (1991). Review of research on teacher participation in school decision making. In G. Grant (Ed.), *Review of research in education* (pp. 225-265). Washington, DC: American Educational Research Association.

Corbett, H.D. (1991). Community influence and school micropolitics: A case example. In J. Blase (Ed.), *The politics of life in schools: Power, conflict, and cooperation* (pp. 120-138). Newbury Park, CA: Sage.

Dreeben, R. (1968). *On what is learned in school*. Reading, MA: Addison-Wesley.

Etzioni, A. (1961). *A comparative analysis of complex organizations*. Glencoe, IL: Free Press.

Etzioni, A. (1975). *A comparative analysis of complex organizations* (rev. ed.). New York: Macmillan.

Foster, W. (1986). *Paradigms and promises: New approaches to educational administration*. Buffalo, NY: Prometheus.

French, J.R., & Raven, B.H. (1968). Bases of social power. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (pp. 259-270). New York: Harper & Row.

Giroux, H. (1981). *Ideology, culture and the process of schooling*. Philadelphia: Temple University Press.

Glaser, B.G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161-179.

Greenfield, W. (1991). The micropolitics of leadership in an urban elementary school. In J. Blase (Ed.), *The politics of life in schools: Power, conflict, and cooperation* (pp. 161-184). Newbury Park, CA: Sage.

Hanson, M. (1976). Beyond the bureaucratic model: A study of power and autonomy in educational decision-making. *Interchange, 7*(1), 27-38.

Hargreaves, A. (1991). Contrived collegiality: The micropolitics of teacher collaboration. In J. Blase (Ed.), *The politics of life in schools: Power, conflict, and cooperation* (pp. 46-72). Newbury Park, CA: Sage.

Henderson, A.H. (1981). *Social power: Social psychological models and theories*. New York: Praeger.

Hollander, E.P. (1978). *Leadership dynamics: A practical guide to effective relationships*. New York: Free Press.

Homans, G.C. (1958). Human behavior as exchange. *American Journal of Sociology, 63*, 507-606.

Hoy, W.K., & Brown, B.L. (1986, April). *Leadership of principals, personal characteristics of teachers, and the professional zone of acceptance of elementary teachers*. Paper presented to the annual meeting of the American Educational Research Association, San Francisco.

Hoyle, E. (1986). *The politics of school management*. London: Hodder & Stoughton.

Imber, M., & Duke, D.L. (1984). Teacher participation in school decision-making: A framework for research. *Journal of Educational Administration, 22*(1), 24-34.

Jacobs, T.O. (1970). *Leadership and exchange in formal organizations*. Alexandria, VA: Human Resources Research Organization.

Johnston, G.S., & Venable, B.P. (1986). A study of teacher loyalty to the principal: Rule administration and hierarchical influence of the principal. *Educational Administration Quarterly, 22*(4), 4-27.

Kleine-Kracht, P., & Wong, K. (1991). When district authority intrudes upon the local school. In J. Blase (Ed.), *The politics of life in schools: Power, conflict, and cooperation* (pp. 96-119). Newbury Park, CA: Sage.

Leithwood, K., & Jantzi, D. (1990, April). *Transformational leadership: How principals can help reform school cultures.* Paper presented to the annual meeting of the American Educational Research Association, Boston.

Lortie, D.C. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

Malen, B., & Ogawa, R.T. (1988). Professional-patron influence on site-based governance councils: A confounding case study. *Educational Evaluation and Policy Analysis, 10*, 251-270.

Marshall, C., & Mitchell, B. (1991). The assumptive worlds of fledgling administrators. *Education and Urban Society, 23*, 396-415.

Mead, G.H. (1934). *Mind, self and society.* Chicago: University of Chicago Press.

Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Mitchell, D.E., & Spady, W.G. (1983). Authority, power, and the legitimation of social control. *Educational Administration Quarterly, 19*(1), 5-33.

National Education Association. (1983). *The national teacher opinion poll* (Research memo). Washington, DC: Author.

National Education Association. (1990). *Estimates of school statistics.* West Haven, CT: Author.

Noblit, G., Berry, B., & Dempsey, V. (1991). Political responses to reform: A comparative case study. *Education and Urban Society, 23*, 379-395.

Optolow, S. (1991). Adolescent peer conflict: Implications for students and for schools. *Education and Urban Society, 23*, 416-441.

Pajak, E., & Blase, J. (1989). The impact of teachers' personal lives on professional role enactment. *American Educational Research Journal, 26*, 283-310.

Pfeffer, J. (1981). *Power in organizations.* Marshfield, MA: Pitman.

Pratt, R. (1988). *The civic imperative: Examining the need for civic education.* New York: Teachers College Press.

Rosenholtz, S.J. (1989). *Teachers' workplace: The social organization of schools.* New York: Longman.

Schlechty, P.C. (1990). *Schools for the twenty-first century: Leadership imperatives for educational reform.* San Francisco: Jossey Bass.

Shakeshaft, C. (1987). *Women in educational administration.* Beverly Hills, CA: Sage.

Sparks, A.C. (1990). Power, domination and resistance in the process of teacher-initiated innovation. *Research Papers in Education, 5*(2), 153-178.

Sparks-Langer, G.M., & Colton, A.B. (1991). Synthesis of research on teachers' reflective thinking. *Educational Leadership, 48*(6), 37-44.

Tartar, C.J., Hoy, W.K., & Bliss, J. (1989). Principal leadership and organizational commitment. *Planning and Changing, 20*(3), 131-140.

Tesch, R. (1988, April). *The contribution of a qualitative method: Phenomenological research.* Paper presented to the annual meeting of the American Educational Research Association, New Orleans.

Thibault, J.W., & Kelley, H.H. (1959). *The social psychology of groups.* New York: Wiley.

Waller, W. (1932). *The sociology of teaching.* New York: Wiley.

Wynne, E.A. (1987, April). *Schools as morally governed institutions.* Paper presented to the annual meeting of the American Educational Research Association, Washington, DC.

Yeakey, C.C. (1987). Critical thought and administrative theory: Conceptual approaches to the study of decision-making. *Planning and Changing, 18*, 23-32.

The Alberta Journal of Educational Research *Vol. XL, No. 1, March 1994, 95-104*

Book Reviews

School Ways: The Planning and Design of America's Schools. Ben Graves. New York: McGraw-Hill, 1993, 237 pages, hardcover, US$39.50.

Reviewed by *George Buck, University of Alberta*

As classroom teachers, school administrators, and teacher educators should we be concerned about the design and construction of school buildings, or should we leave such matters to architects and others hailed as design experts? It has been contended by some that where school takes place should be a minimal consideration in education. Indeed, it may be argued, hyperbolically at least, that all that is required for a school is a log with a pupil at one end and a teacher at the other with little regard for the length, condition or suitability of the log or its context (Lewis, 1937). Such a learning environment leaves much to be desired, especially in winter. On the other hand, excessive concern about the appearance or image of the school plant at the expense of the individuals who must function within its precincts, namely the pupils and teachers, can be equally detrimental to education. Although books describing existing school architecture and works that suggest how to improve the design and construction of new school buildings appear to have been much more common in the past than at present (American Association of School Administrators, 1949; Building Research Institute, 1962; Caudill, 1954; Harrison & Dobbin, 1931), Ben Graves' book *School Ways: The Planning and Design of America's Schools* seeks not only to provide a current summary of successful school designs, but also to recommend directions for the planning of new school buildings. Although the book's title suggests that the subject matter is relevant only to the United States, many of the designs and concepts discussed in the book have been used, and continue to be used in some instances, in areas of Canada, so the book is relevant to the Canadian scene.

The work begins with an identification and elucidation of some of the indicators and factors used in forecasting the need for new school buildings. The next chapter presents an all-too-brief summary of selected developments in United States school architecture from 1800 to the present. Some of the current issues in school architecture are then addressed followed by descriptions of several planned projects and possible designs of future school buildings. The remaining chapters provide information and suggestions on how to engender and implement the planning process of school design and construction so that the outcomes are successful. A helpful listing of resource materials plus names and addresses of institutions relevant to school building design and construction conclude the work.

Between all but the first and second chapters are *Portfolios*. These are photographs, plans, and descriptions of selected innovative school designs, deemed successful by the author, and that are representative of the subject matter in the preceding chapter. The photographs, several in color, generally complement the descriptions and plans well. For example, the first portfolio, concerned with earlier innovative designs that have had extensive influence on sub-

sequent construction, contains one of the most informative and clear descriptions of the Crow Island school in Winnetka, Illinois to appear in any book on school architecture known to the reviewer.

Unlike some earlier books on school architecture, Graves' work does not focus exclusively, or even primarily, on large urban schools. Consideration is given to smaller centers and rural settings, including one portfolio describing recent designs of one-room schools used in remote areas. The text is written so that it may be understood by those who do not have extensive knowledge of architecture and architectural terms. Thus the book is accessible to almost anyone involved with or concerned with the design and development of school architecture.

In no small way, however, the introduction by William Brubaker detracts from the perceived usefulness of the work. Although some people like to predict future developments and trends and the speed at which they will occur, especially developments that are envisaged as revolutionizing or altering education radically, such predictions have been and continue to be inaccurate, and their inclusion encourages readers to question the credibility of the work. For instance, Brubaker mentions that some schools of the future may extend up to 12 stories and that it is possible for, "a high school to be in an office building" (p. 6). Although it is possible that a few schools of this configuration may be erected in extremely large urban centers (Graves describes the new 10-story Stuyvesant school in New York City), this prediction is little changed from a similar one made by Harrison (1931) 62 years earlier, where he envisaged skyscraper schools complete with escalators and integration with offices and shop. Given that such school buildings have not become common in spite of two predictions made 62 years apart, one might conclude that there is a fundamental difference in what are considered important design features between some architects and those in education responsible for the selection of school building designs.

A more serious criticism concerns Graves' understanding and criticism of particular aspects of education. Throughout the text, Graves refers to *educators*, yet he does not define what he means. This is not a trivial point, because he mentions in more than one place that it is imperative that there be cooperation between educators, architects, manufacturers, and students to ensure proper school design. Although consideration is given to input from parents, pupils, school administrators, consultants, and board members, one body of educators, teachers, is conspicuous by its absence. Although the average teacher may not know much about architecture, teachers do work in particular schools most of the time, unlike most administrators, consultants, and board members. Teachers, therefore, possess some idea of what aspects of design are suitable and which are not suitable in a school. Moreover, many school administrators, board members, and even government officials began as classroom teachers. Washburne (Washburne & Marland, 1963), the superintendent associated with the planning and construction of the Crow Island school, states that the creation of this school and the enduring success of its design came about because all parties, including teachers, were involved in its design. Whenever Graves mentions teachers it is usually in disparaging tones. Describing the open-plan (or open-area) school concept popular for a time during the 1950s and 1960s,

Graves claims that this architectural and educational concept failed primarily because teachers were unwilling to adapt to this innovative idea that embodied the principles of the old one-room schools (p. 30). This sentiment contradicts both the general tenor of the book, that it is only through collaboration that successful school design results, and Graves' earlier claim that, "Fads such as ... windowless classrooms, stock plans and classes on buses ... have all come and gone" (pp. 27-28). If windowless classrooms were a fad, then it follows that open-plan schools were a fad too, because their design and general enthusiasm for this type of school prevailed for only a brief period throughout North America. In spite of Graves' claim that open-area schools were similar in principle to the older one-room schools, it was unusual for more than one teacher to be teaching in a one-room school at a time; although students in one grade could hear the lessons for the other grades, there was only one teacher present. The problem of sound in open-plan schools was recognized and noted by some architects before such schools gained popularity and enthusiastic promotion. Caudill (1954) mentions two alternatives available for resolving this problem, "new methods for controlling sound in these open spaces, or (2) we condition ourselves and our pupils to be less sensitive to disturbing noises" (p. 87). Graves' criticism of teachers' role in the demise of open-area design schools is perplexing in another plane, because elsewhere in the book he expresses no concern that a particular group of individuals in a community, who were not professional educators, objected to the planned use of pink brick in a school (p. 199).

Graves does describe many examples of successful modern designs of school buildings, including descriptions of materials used. We are offered hope that future school building designs need not be uninspired and user unfriendly, when Graves states, "the days of the hermetically sealed institution, isolated from its surroundings are over" (p. 73). Budgetary limitations common in many school districts currently are addressed, as examples are presented of attractive and functional schools that were comparatively inexpensive to create. The author notes that there has been a gradual return to the tradition of using natural light in the classroom rather than using artificial lighting exclusively, and many of the schools described in the book reflect this trend. Unfortunately, there is no mention as to which specific type of artificial lighting (fluorescent, sodium vapor, incandescent, halogen) should be used. It is possible, therefore, for planners to select artificial lighting on the basis of cost only, ignoring or neglecting possible health factors associated with specific artificial light sources.

Portable buildings are mentioned as being sometimes required to augment an existing school building, yet neither plans of particularly successful examples nor suggestions for the design of such buildings are provided. This omission seems peculiar, because Graves reports elsewhere in the text that in one location residents opposed an attempt to replace an old school building with trailer modules. We are left wondering whether prefabricated portable structures such as those produced by ATCO are suitable at all, or whether custom designed and constructed portables are preferable.

Graves mentions accommodation of physically disabled people in modern school buildings several times, a most important consideration given the cur-

rent direction of integration of physically handicapped students. Unfortunately, apart from mentioning the use of ramps, little else is described, and the reader is left wondering if there are any innovative designs or techniques that could be shared.

Graves notes that some aspects of school design have become obsolete primarily because individuals in education have neglected them. One important example is the design and improvement of school furniture. Several schools are described as possessing computer rooms where the apparatus is placed on tables of a fixed height. One can envisage problems if the computers are used both by grade 1 and grade 6 students and wheelchair users. Although it is true that detachable keyboards accommodate some individual differences, it is preferable for the computer's monitor and other peripherals such as mice to be at a height convenient to the user. Graves suggests that effort should be directed toward "requiring furniture that will adjust to the user's size and other needs, such as an understanding that left-handed people are uncomfortable with right-handed furniture" (p. 202), as he notes was done much earlier in the design of the furniture for the Crow Island school. Indeed, chapter two of Dewey (1900) and the first issues of the *School Board Journal* (1891-1894) reveal that considerable thought was given to the design of school furniture. It is time to resurrect such interest.

One wonders if developments in school architecture in other countries has influenced current United States designs. Although Graves does not mention such influence, McLeod (1962) compares the British *CLASP* system of school design with contemporary processes in the United States. Did any such foreign developments effect change to North American designs, or are current designs based solely on ideas originating in the United States?

Although several criticisms have been noted in *School Ways*, I recommend it nevertheless as a suitable resource for those involved in the planning and design of new school buildings. Moreover, it would be an excellent textbook for a course or a portion of a course to acquaint preservice teachers with some of the history, planning, designs, considerations, and problems associated with modern school architecture. If we wish to improve all aspects of education, then we must ensure that the teachers we graduate possess knowledge beyond pedagogy and curriculum. Thus educators may be better prepared to guide and assist architects and furniture designers in the planning and construction of appropriate buildings and infrastructure.

References

American Association of School Administrators. (1949). *American school buildings*. Washington, DC: Author.

Building Research Institute. (1962). *School building research*. Washington, DC: Author.

Caudill, W.W. (1954). *Toward better school design*. New York: Dodge.

Dewey, J. (1900). *The school and society*. Chicago: University of Chicago Press.

Harrison, W.K. (1931). The design of the modern school. In W. Harrison & C. Dobbin (Eds.), *School buildings of today and tomorrow*. New York: Architectural Book.

Harrison, W.K., & Dobbin, C.E. (1931). *School buildings of today and tomorrow*. New York: Architectural Book.

Lewis, C.D. (1937). *The rural community and its schools*. New York: American Book.

McLeod, J.W. (1962). Materials and construction needs for future educational facilities. In *School building research*. Washington, DC: Building Research Institute.

Washburne, C.W. & Marland, S.P. (1963). *The history and significance of an educational experiment.* Englewood Cliffs, NJ: Prentice-Hall.

George Buck is a lecturer in the Department of Educational Psychology. He began his career as a teacher and includes school building design and its relationship to learning and instructional theory among his areas of interest.

Testing Testing: Social Consequences of the Examined Life. F. Allan Hanson. Berkeley, CA: University of California Press, 1993, 378 pages, Can$38.95.

Reviewed by *Birendra K. Sinha, University of Alberta*

This book is a scholarly treatise of intentional as well as unintended social consequences of testing in the contemporary United States. The author, an anthropologist, has presented a detailed analysis of the many facets of the testing enterprise from a sociocultural perspective. Two main themes have been expounded in this book. First, the tests "create" individual differences in respect to specified personal characteristics. Second, testing is an instrument of power exercised by a select group of persons or institutions over ordinary citizens. In general, Hanson argues that the ostensible purpose of testing—assessment of the existing personal characteristics and prediction of future behaviors—is misleading in that it obfuscates the altered relationships between test givers and test takers. As a consequence, new but subtle social realities are produced that are antithetical to the goals of a democratic society where "equal opportunity" and "unequal endowments" are attempted to be reconciled.

There are 10 chapters in the book including Introduction and Conclusion. The main chapters are divided into two parts: Authenticity Tests and Qualifying Tests. The five chapters on Authenticity Tests in Part I discuss the prescientific history and modern practice of determining innocence, guilt, honesty, and truthfulness of individuals and groups concerning culturally important behaviors (e.g., lie detection and drug testing). The Qualifying Tests in Part II refer to ability tests, such as achievement, aptitude, and intelligence for the purpose of selecting and promoting qualified applicants in academia, industry, and the military. These qualifying tests are also used for vocational guidance and career counseling to ensure better fit between the individual and society.

As a social scientist Hanson clearly specifies his basic premise and assumptions in the introductory chapter. He is convincing in emphasizing that the extreme pervasiveness of testing in the United States stems from the philosophical underpinning of "positivism," the fundamental tenet of which is that science can provide better understanding and discover laws of human behavior. Thus not only those engaged in the testing enterprise but also the political leaders and the general public have developed unshakable faith in the usefulness of tests. Hanson is at his best when discussing the early history of "authenticity testing" where he adduces many historical events to support his arguments that tests are simply representations of culturally sanctioned reality. Trials by battle, ordeal, torture, and similar practices in preindustrial Europe could be construed as precursors of modern tests where application of power by authorized institutions or their representatives to an individual is the common theme. The author devotes considerable space to lie detection and associated polygraph testing. Through wit, logic, and empirical evidence,

Hanson demonstrates the unscientific nature of lie detection tests and grave injustice perpetrated on powerless individuals. The author does acknowledge that evidence of guilt or innocence based on polygraph testing is not admissible in courts of law and that legislative measures have been taken in the United States to control or prohibit lie detection tests. However, it is interesting to read that polygraph testing still enjoys a certain degree of acceptance among lawyers, judges, police, and employers. For example, as recently as 1991, Anita Hill attempted to establish her credibility by passing a polygraph test during the United States Senate Judiciary Committee hearings regarding her sexual harassment charge against Judge Clarence Thomas. Because of the legal restrictions on polygraph testing, a number of paper-and-pencil "integrity tests" (i.e., honesty tests) have been developed for use with employees in industrial settings. Although these tests are not supported by reputable psychometricians, they have virtually replaced polygraph testing in personnel work. The author rightly cautions the reader to be skeptical of the scientific merits of these written integrity tests.

Hanson raises similar philosophical and scientific issues in his discussion of drug testing with urinalysis in schools, universities, workplace, and sports. Although it is important to recognize the danger of drug abuse, the author draws attention to the cost of routine drug testing and accuracy of the findings. Undoubtedly significant numbers of false positives have destroyed the dreams and careers of many young persons in North America. Notwithstanding the media presentations, drug use is on decline in the United States. Then why are we so concerned? Hanson suggests that institutionalized drug testing is aimed at social control through mostly negative sanctions (e.g., job loss or dishonorable discharge). In the "disciplinary technology of power," drug testing is an attempt at mind control. The test givers are culturally sanctioned to exercise power over test takers by determining the type and method of information extraction.

In the section on Qualifying Tests, the author examines selection tests in the context of control and power. One chapter presents a fascinating account of the history of ability testing not generally discussed in standard textbooks on psychological and educational testing. Starting from the Chinese Civil Service Examination in the 7th century to the more recent Western attempts through physiological (e.g., phrenology), oral, and written examinations, the goal has been to select persons best suited to various positions in the society. But as Hanson points out, all such tests are biased in favor of the class in power, and as such are not truly objective. The superior performance of the white, middle-class, anglo-saxon test takers are explained in terms of their superior genetic endowment. Similarly, the rise of vocational and interest testing is primarily for the masses. Test givers have never suggested that one of the qualifications for seeking political office, for example, should be acceptable performance on selected aptitude and interest tests. Familiar concepts, such as intelligence, eugenics, race, class, and self-esteem are critically evaluated by Hanson to show that they are infinitely more complex issues than assumed by test givers.

Criticism of psychological, educational, and physiological testing is not new. Many books and articles have been written during the last 40 years emphasizing the harmful consequences of ability and personality testing. What

is new in the book under review is that the author has successfully demonstrated the historical continuity of the testing enterprise for the purpose of exercising social control. Furthermore, the two major constituents of the testing enterprise—test givers and test takers—are working at cross purposes. Advances in the technology of testing have not changed the real purpose of testing from that in the prescientific era of trials by battle and ordeal. Thus modern testing is neither objective nor democratic as claimed by test givers.

The author is commended for dealing with a complex issue of contemporary society with relevant facts and critical analysis. However, there are several lacunae in Hanson's exposition of testing in the United States. First, he does not give adequate consideration to the consequences of not using tests, particularly qualifying tests. Deeply entrenched sociopolitical and cultural biases can be overcome through higher performance on tests as has happened in case of admission to Canadian medical schools. A large number of ethnic minorities are entering professional programs in Canada because of their superior performance on the standardized tests in spite of the fact that some ethnic groups are least liked by the majority. Second, the author virtually ignores the psychometric advances in modern testing. Hanson's definition of a test as a "representational technique" is too vague compared with the generally accepted definition that a test is a systematic and quantitative observation of a sample of behavior for prediction purposes. Because prediction is the ultimate aim of testing, the contribution of decision theory and multiple regression methods in improving prediction regardless of the "power motive" of test givers deserves serious discussion. There are certainly ways in which the interests of the test takers can be safeguarded in a democratic society.

Birenda K. Sinha is an associate professor in the Departments of Psychology and Psychiatry and Director of the Psychometric Laboratory in the Psychology Department. He specializes in experimental and clinical psychopathology including psychometrics.

Pedagogue's Progress. Fred S. Keller. Lawrence, KA: TRI Publications, 99 pages.

Reviewed by *Terry W. Belke, University of Alberta*

Pedagogue's Progress is an autobiographical account of Fred Keller's experiences that documents how traditional methods of instruction failed him both as a student and as a teacher and how this failure motivated his quest for alternative methods of instruction that eventually led to the development of a personalized system of instruction (PSI). The book is written in the informal voice of personal recollections. Chapters are short narrative accounts (2-4 pages) organized in a chronological progression starting with Keller as a youth searching for an occupation and ending with Keller as a retired academic making predictions about the future of education.

In the introduction, Keller states that this book is written for educated lay people—students both past and present. Specifically, he wishes to address "alumni and alumnae who are dismayed to see their alma mater with feet of crumbling clay" (p. 1) as well as students currently in the educational system

who feel something is wrong with their schools, their teachers, or themselves. According to Keller, the book is a story of the failure of a student, a teacher, and the educational system of which they were a part. From this failure arose his dismay with traditional methods of instruction, which motivated the development of a personalized system of instruction (PSI) based on principles of operant conditioning.

The book is divided into five sections. Chapters 1-4 offer recollections of early studies, early occupations, and early years of teaching. Chapters 5-6 document Keller's first attempt at developing a new method of instruction to train students in the use of telegraphy for the military. Chapters 7-11 deal with Keller's efforts to improve his teaching through the application of principles of reinforcement. Chapters 12-21 deal with the application of PSI in a formal educational system in Brazil. Finally, Chapters 22-25 recount the origin and growth of PSI as an alternative to group instruction.

The first section begins with Keller as a high school dropout exploring different occupations and settling on a job as a messenger boy for Western Union. During this time, Keller developed a fascination with telegraphy. Following service in the army during World War I, Keller took advantage of military-sponsored funds to complete his education. In this early period Keller found the motivation to complete his education after reading Watson's (1919) *Psychology from the Standpoint of a Behaviorist*. On completing his bachelor's degree, Keller was accepted for graduate study at Harvard where, Keller states, the most important thing he learned was from his fellow graduate student Burrhus Frederic Skinner. His observation of one of Skinner's rats in an operant conditioning chamber learning to respond on a lever for food left a lasting impression.

Keller obtained a position as an instructor at Tufts college, where his attempts at teaching using the traditional lecture method began. The basis for this method, as he later found out, was *The Great Didactic* by Johann Comenius (Jan Amos Komensky) in 1638. However, despite adhering to Komensky's prescriptions, Keller judged himself a failure as a teacher. As he points out, the prescriptions of Komensky do not refer to the effects of the method on the scholar. "In spite of my efforts with Komensky's methods, no more than ten percent of all my pupils accomplished what I felt they should. At least one half of them did no better than I had done in college" (p. 14) To Keller, the virtues of the traditional method were merely to inspire listeners or showcase the talents of the speaker, but *not* to produce effective learning in scholars.

The second section documents Keller's first effort at developing a new method of instruction. Keller recalled that rat in the operant conditioning chamber pressing a lever in the presence of a light and receiving immediate feedback. Based on this observation, he devised a code-voice method to train telegraphers in which subjects heard the code, recorded the letter that the code represented on a special code sheet, and then heard the letter that the code represented over a speaker. The system provided immediate feedback to subjects about the correctness of their responses and allowed them to monitor their own progress and to progress at their own pace. The system was successful. Keller was hired by the military to train telegraphers; however, an evaluation team of civilian educators failed to see the advantages of Keller's method.

Section three describes Keller's return to academia, still dissatisfied with the traditional method of instruction and motivated to generalize principles of conditioning to the instruction of an introductory course in psychology. He designed the course to make instruction more systematic and allow students to observe principles of conditioning directly through experimentation. Despite the success of these changes, Keller claims that he was still not a successful teacher. His feeling of failure arose from the distribution of grades in his courses, which showed a normal distribution of grades in which the proportion of students at each level did not vary. Did this reflect something about the students, or was it a product of the teacher grading on the curve? Keller was sensitive to the expressions on the faces of the students when they learned of the outcomes of their efforts. Furthermore, Keller compared himself with his eminent colleagues and found himself wanting.

> I possessed a certain narrowness of outlook, a certain inability to see all sides of every question, and a tendency to act in the absence of exhaustive preparation. Added to these was a blind enthusiasm for every project that I undertook; and, to top it all, a strong dislike for engaging in open debate. These are not the characteristics of a professor, let alone a scientist or scholar. They are more nearly the traits of a fanatic, a promotor, or a missionary; and perhaps my colleagues saw this. (pp. 30-31)

This marked the depth of Keller's despair. He felt like an outsider among his colleagues and questioned his qualifications as an educator.

In the chapters that follow, Keller reviews the collection of facts and principles that comprise reinforcement theory and a programmed system of instruction. Following this, he recounts an occasion when Skinner visited the classroom of his daughter's teacher. Skinner observed that the students were motivated to work by escape from the threat of punishment, that delays to reinforcement for correct behavior were long, that large blocks of behavior were reinforced rather than each correct behavior, and that there was a general lack of reinforcement in the learning context. Based on these observations, Skinner devised a way to generalize laboratory laws of conditioning to the classroom that led to the development of a system of programmed instruction as well as the programmed textbook and teaching machine. In reading about Skinner's work, Keller found the solution to his despair.

The fourth section documents how PSI came to be developed in a formal educational setting in Brazil. Keller moonlighted by teaching courses in University Extension. In one of his extension courses, he met a young woman by the name of Myrthes Rodriguez do Prado, who left in the middle of his class. Although he never saw her again, their meeting affected the course of his life. In 1959 he received a letter from the Dean of the Faculty of Philosophy, Sciences, and Letters at the University of São Paulo inviting him to come to the university to teach experimental and comparative psychology. The letter was signed by his former student. The subsequent year in São Paulo challenged Keller's stereotypes and showed that some props he considered to be essential to teaching were unnecessary. Keller left having lit the torch of reinforcement theory in Brazil.

In Chapter 18, Keller describes his dream course based on programmed instruction designed to teach reinforcement theory. It is a course with lectures,

demonstrations, discussions, laboratory hours, and homework. Lectures and demonstrations would be infrequent and inspirational. Attendance at lectures and demonstrations would be optional. Each student would have access to a room or cubicle well stocked to run experiments. Students would write a report on each experiment and answer a question on the assigned readings. The role of the teacher would be to program the instruction for each student and redesign the course based on each student's performance. When all course requirements were met the course would end. There would be no course grade, no final examination, no reward for speed of attainment, and no punishment for delay. Keller concludes this course is unlikely to be embraced in North America where the traditional method is entrenched; however, it could happen elsewhere; his dream course was implemented in Brazil and judged to be a success.

In the last section, Keller provides a prescription for a personalized system of instruction (PSI) for any teacher who may be inclined to follow in his footsteps. His chapter on individualized instruction is analogous to Komensky's prescription for the group instruction using the lecture method. In the final chapter, Keller looks back in retirement over the road he has travelled and makes a bold prediction about the future role of the teacher. PSI made his final experiences with teaching rewarding; however, he concedes that his ultimate failure was in not convincing more educators that good teaching is more a matter of arranging conditions for happy and effective learning than "show and tell." Keller looks into his crystal ball and foretells the passing of teachers as we know them and the demise of group instruction.

Pedagogue's Progress persuasively makes the point that the education system has failed its constituents and that more effective methods of instruction based on principles of conditioning are available. It carries a message similar to that made by Skinner (1987) in *The Shame of American Education*, but makes the point in a different way. At the very least the book leads one to challenge some of the fundamental assumptions about how to best instruct students. Keller's recollections speak to all who have experienced dissatisfaction with the accomplishments of their current methods of teaching and who seek a better way to reach their students.

References

Skinner, B.F. (1987). The shame of American education. In B.F. Skinner (Ed.), *Upon further reflection*. Englewood Cliffs, NJ: Prentice-Hall.

Watson, J.B. (1919). *Psychology from the standpoint of a behaviorist*. Philadelphia, PA: Lippincott.

Terry Belke recently completed a doctorate in psychology at Harvard and holds a postdoctoral fellowship at the University of Alberta.

Index, Volume XXXIX, 1993

Ajar, D. *Formative Assessment in a Classroom Setting: From Practice to Computer Innovations*, No. 1, 111-125.

Allston, D. *Effective Teaching Does Make a Difference*, No. 2, 191-203.

Anderson, J.O. *A Time and Place for Observations: Talking with Primary Teachers About Classroom Assessment*, No. 3, 363-374.

Baby, A. *Educational Research in Quebec*, No. 1, 7-21.

Baby, A. *Integration of Youth into the Labor Market: Some Theoretical Considerations on a Structural/Conjunctural Issue*, No. 1, 97-109.

Bacon, L. *Socioconstructivist Viewpoint on the Use of Symbolism in Mathematics Education*, No. 1, 41-58.

Baillargeon, M. *One Child, Many Environments: Continuity or Discontinuity in Kindergarten and School-based Day Care Programs?* No. 1, 127-142.

Bednarz, N. *Socioconstructivist Viewpoint on the Use of Symbolism in Mathematics Education*, No. 1, 41-58.

Bergeron, J.C. *An Epistemological Study of the Child's Construction of Early Arithmetic*, No. 1, 23-40.

Betsalel-Presser, R. *One Child, Many Environments: Continuity or Discontinuity in Kindergarten and School-based Day Care Programs?* No. 1, 127-142.

Brassard, A. *Conceptions of Organizational Effectiveness Revisited*, No. 1, 143-162.

Buck, G.H. *Faculty of Education, University of Alberta: A 30-year Gestation*, No. 3, 319-334.

Caron, A. *Learning to Think: Philosophical Instruction and Reflective Educational Practice*, No. 1, 73-83.

Coleman, P. *The Politics of Excellence: Trustee Leadership and School District Ethos*, No. 4, 449-475.

Cousins, J.B. *Teachers' Evaluation of Correlational Reasoning Skills*, No. 3, 297-317.

Daignault, J. *On Writing Teaching*, No. 1, 85-95.

Daniel, M.-F. *Learning to Think: Philosophical Instruction and Reflective Educational Practice*, No. 1, 73-83.

Dassa, C. *Formative Assessment in a Classroom Setting: From Practice to Computer Innovations*, No. 1, 111-125.

Derwing, T. *Literacy Learning of Adults in a Bilingual ESL Classroom*, No. 4, 393-406.

Désautels, J. *"The Map is the Territory!" The Viewpoints of Biology Students on the Theory of Evolution*, No. 1, 59-72.

Désautels, J. *Editorial.* No. 1, 1-7.

Dufour-Janvier, B. *Socioconstructivist Viewpoint on the Use of Symbolism in Mathematics Education*, No. 1, 41-58.

Dufresne-Tassé, C. *Editorial.* No. 1, 1-7.

Duquette, C. *A School-based Teacher Education Program: Perceptions and Attitudes*, No. 4, 419-432.

Ellefson, I. *Academic Outcomes and Behavior, No. 2, 235-243.*

Geddert, P. *Student Success Through Outcome-based Education*, No. 2, 205-215.

Herscovics, N. *An Epistemological Study of the Child's Construction of Early Arithmetic*, No. 1, 23-40.

Himsl, R. *Signs of Learning in the Affective Domain*, No. 2, 257-272.

Joncas, M. *One Child, Many Environments: Continuity or Discontinuity in Kindergarten and School-based Day Care Programs?* No. 1, 127-142.

Kulmatycki, M.B. *A Comparison of Principals' Perceptions About Their Leadership Roles in Catholic and Non-Catholic Schools*, No. 3, 375-392.

Laflamme, C. *Integration of Youth into the Labor Market: Some Theoretical Considerations on a Structural/Conjunctural Issue*, No. 1, 97-109.
Lambert, E. *Signs of Learning in the Affective Domain*, No. 2, 257-272.
Larochelle, M. *"The Map is the Territory!" The Viewpoints of Biology Students on the Theory of Evolution*, No. 1, 59-72.
LaRocque, L. *The Politics of Excellence: Trustee Leadership and School District Ethos*, No. 4, 449-475.
Larouche, H. *One Child, Many Environments: Continuity or Discontinuity in Kindergarten and School-based Day Care Programs?* No. 1, 127-142.
Lebuis, P. *Learning to Think: Philosophical Instruction and Reflective Educational Practice*, No. 1, 73-83.
Levin, B. *Financing Education in the Northwest Territories: Reconciling Competing Values*, No. 3, 347-361.
MacDonald, C.J. *Coping with Stress during the Teaching Practicum: The Student Teacher's Perspective*, No. 4, 407-418.
Maguire, T.O. *Some Thoughts on the Alberta Version of Educational Quality Indicators*, No. 2, 275-277.
Malicky, G. *Literacy Learning of Adults in a Bilingual ESL Classroom*, No. 4, 393-406.
McEwen, N. *Editorial*, No. 2, 165-165.
McEwen, N. *Educational Quality Indicators*, No. 2, 167-177.
McLean, L.D. *Bold Projects, Prudent Measures*, No. 2, 279-282.
Mestinsek, R. *District and School Profiles for Quality Education, No. 2, 179-189.*
Montgomerie, T.C. *A Comparison of Principals' Perceptions About Their Leadership Roles in Catholic and Non-Catholic Schools*, No. 3, 375-392.
Nicholson, D.J. *A Time and Place for Observations: Talking with Primary Teachers About Classroom Assessment*, No. 3, 363-374.
Owen, M. *Rev. Dr. George W. Kerby: Evangelist for the Home and School*, No. 4, 477-493.
Pisichko, K. *Manitoba's GSE: An Attempt at Resolving the Fiscal Equity Dilemma*, No. 3, 335-346.
Poirier, L. *Socioconstructivist Viewpoint on the Use of Symbolism in Mathematics Education*, No. 1, 41-58.
Prentice, M. *Teachers' Evaluation of Correlational Reasoning Skills*, No. 3, 297-317.
Ralph, E.G. *Sensitive, Sensible Practicum Supervision: A Contextual Application in Saskatchewan*, No. 3, 283-296.
Ross, J.A. *Teachers' Evaluation of Correlational Reasoning Skills*, No. 3, 297-317.
Rymhs, R. *Effective Teaching Does Make a Difference*, No. 2, 191-203.
Schleifer, M. *Learning to Think: Philosophical Instruction and Reflective Educational Practice*, No. 1, 73-83.
Schulz, L. *Effective Teaching Does Make a Difference*, No. 2, 191-203.
Sereda, J. *Educational Quality Indicators in Art and Mathematics*, No. 2, 217-234.
Vázquez-Abad, J. *Formative Assessment in a Classroom Setting: From Practice to Computer Innovations*, No. 1, 111-125.
Willard, L.A. *Identifying and Measuring Desirable Student Social Outcomes*, No. 2, 245-256.
Young, J.H. *Principals and Collaborative Curriculum Development: Does Gender Make a Difference?* No. 4, 433-448.
Zaïm-Idrissi, K. *"The Map is the Territory!" The Viewpoints of Biology Students on the Theory of Evolution*, No. 1, 59-72.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes;* any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
4-116 Education North
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1436

40

ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XL NUMBER 2 JUNE 1994

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

The Alberta Journal of Educational Research

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
University of Alberta

George Buck
University of Alberta

James Chapman
Massey University

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Linda LaRocque
University of Alberta

Stewart J.H. McCann
University College of Cape Breton

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Sharon Robertson
University of Calgary

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Graham Fishburne
Robert H. Short
Fern Snart

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Publication Services
4-116 Education Centre North
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0390

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XL, Number 2, June 1994

Robert J. Wilson *John R. Kirby*	**105**	Introduction: Special Issue on Cognition and Assessment
Thomas Maguire *John Hattie* *Brian Haig*	**109**	Construct Validity and Achievement Assessment
Gay L. Bisanz *Jeffrey Bisanz*	**127**	Studying the Development of Academic Skills in the 1990s: Implications for Assessment
John R. Kirby *Rosamund A. Woodhouse*	**147**	Measuring and Predicting Depth of Processing in Learning
Philip Nagy	**163**	Cognitive Assessment of Solutions to an Ill-structured Problem Using Forced Classification Analysis
Philip H. Winne *Lorene Gupta* *John C. Nesbit*	**177**	Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics
W. Todd Rogers *David J. Bateson*	**195**	Verification of a Model of Test-taking Behavior of High School Seniors
Bikkar S. Randhawa	**213**	Theory, Research, and Assessment of Mathematical Problem Solving
David Bateson	**233**	Psychometric and Philosophic Problems in "Authentic" Assessment: Performance Tasks and Portfolios
Dan G. Bachor *John O. Anderson* *John Walsh* *Walter Muir*	**247**	Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 105-108*

Robert J. Wilson

and

John R. Kirby
Queen's University

Introduction: Special Issue on Cognition and Assessment

The educational field of measurement has changed dramatically in recent years. Teachers are unenthusiastic about the value of large-scale, centrally designed testing, perceiving it to be less than relevant to classroom learning. Administrators and parents worry about some of the methods teachers have adopted for assessing students' success. As the wider community too has begun to challenge the effectiveness of the education system in general, concerns with measuring learning have taken on a political urgency. The time is ripe to reconsider the goals and techniques of educational assessment.

In measurement terms, these issues concern validity, but validity in measuring achievement outcomes is becoming more elusive even as more attention is being paid to it. As some investigators concentrate on the contexts in which assessment takes place and the representativeness of tasks used in those contexts, others concern themselves with the adequacy of any tasks to measure cognitive operations assumed to be the mental resources that create achievement outcomes.

Central to all these interests are the processes invoked to respond to achievement tasks. For some measurement specialists, Bloom's Taxonomy has been the beginning and end of the consideration of processes, although the taxonomy itself is dated and superseded by much recent work in cognition. Conversely, some work in cognition has often treated validity concerns such as construct span and representativeness inadequately and disinterestedly. Both cognition and measurement stand to benefit from a consideration of each other's problems and solutions (Baker, O'Neil & Linn, 1993; Wittrock & Baker, 1991).

Two previous meetings of Canadian researchers interested in measurement issues (Anderson, 1990; Bateson, 1992) led the participants into a concern with the nature of achievement, cognitive processes, and the ways these are assessed. A third meeting, it was decided, should exploit Cognition and Assessment as a theme, and papers were invited from researchers active in both fields. This special issue of the *Alberta Journal of Educational Research* is a result of their efforts.

Robert Wilson is a professor of measurement and evaluation. His research interests are in large-scale assessment and classroom processes in evaluating student learning.
John Kirby is a professor of educational psychology. His research interests include individual differences in cognitive processes, learning disabilities, and study processes.

Measurement and Psychology

The fields of measurement and psychology have been closely allied throughout their histories. In many universities they inhabit the same department. Because educational measurement has been concerned primarily with assessing students' knowledge, skills, and problem solving; and cognitive psychology has been concerned with how knowledge, skills, and problem solving function are represented in the brain, it has seemed sensible to consider them as mutually supporting and enriching.

Until recently educational measurement has been closely aligned with psychometrics, a tradition in psychology that is more mathematical than cognitive. Although psychometrics (literally measurement of the mind) should be cognitive in orientation, it began its development during the behaviorist domination of learning theory. Lacking guidance on appropriate cognitive constructs to pursue, psychometricians focused instead on the development of mathematical techniques such as factor analysis for analyzing test results, thus shifting attention away from the need to conceptualize about the cognitive nature of the mathematical constructs. The result was a technology of assessment that grew out of a test theory that lacked a basis in psychological theory, but that nevertheless came to be a distinct scientific discipline in education and psychology (Cronbach, 1957; Ferguson, 1954).

Cognitive psychology should be a source for the constructs that underlie educational achievement, and may be able to shed some light on how to measure these constructs (Biggs & Collis, 1982). However, cognitive psychology had its roots in experimental psychology, the domain of operational definitions. To illustrate: when two experimental conditions are contrasted for their effect on concept identification, great care is taken in constructing the stimulus lists, but little thought is devoted to the representativeness of the content, the instructional method, or the particular measure of concept learning.

Validity issues are inescapable, but cognitive psychologists seem on safe ground in such studies—until they begin to generalize. Once inferences are drawn from this work about how to teach students real concepts, the concern for validity grows. The issue becomes even more important when cognitive psychologists turn their attention to more complex constructs such as those in theories of mental ability (e.g., Carroll, 1976; Das, Kirby, & Jarman, 1975; Sternberg, 1977). Although it is unlikely that cognitive psychology has any simple solutions for educational assessment, it may be able to contribute to finding those solutions through its work on modeling underlying cognitive processes that give rise to observed measurements.

In this issue we see a number of cognitive psychologists approaching the assessment of educational achievement, employing the techniques of cognitive psychology, but with concern for validity: Bisanz and Bisanz address the cognitive processes involved in reading and arithmetic; Kirby and Woodhouse consider methods for predicting and measuring depth of processing in learning; Nagy presents a technique for analyzing solutions to ill-structured problems that might help describe learners' cognitive processes; Randhawa focuses on the processes of mathematical problem solving; and Winne, Gupta, and Nesbit illustrate techniques for measuring students' learning strategies.

Authentic Assessment

Those who work in educational assessment did not need cognitive psychology to begin to change. There has been growing disenchantment among educators over traditional forms of educational assessment. Among other concerns has been the sense that centrally planned, standard-format tests do not tap what teachers are trying to teach, or even what curricula indicate they should be teaching. New assessment methods have been devised under the collective label "authentic assessment" (Wiggins, 1989). Bateson in this issue provides a useful overview of how this movement has both annoyed and inspired measurement practitioners, especially in the field of large-scale assessment.

Validity (of which "authenticity" is but one aspect) is specifically addressed in this issue in the articles by Maguire, Hattie, and Haig, and Bachor, Anderson, Walsh, and Muir. Among other issues, Maguire et al. discuss whether a crucial aspect of a measure's validity is the use to which the information is put. They emphasize the need for conceptual links between the construct and its indicator on the one hand, and the indicator and any score on the other. Bachor et al. are also interested in the use of measures. Their approach, however, is to delineate which aspects of tasks are more or less useful for common assessment needs in classrooms. Both of these articles expand the context in which the validity of educational assessment is considered.

Final Words

The articles in this special issue revolve around one central issue: how can cognitive psychology and measurement practice work together to enhance the valid assessment of learning? We believe that these articles contribute to a deepening of the debate about the purposes and methods of educational assessment and provide direction for future work.

We begin with the Maguire, Hattie, and Haig article which raises many of the fundamental issues regarding validity and assessment. The second article, by Bisanz and Bisanz, reviews recent research in cognitive psychology that is directly relevant to the assessment of children's learning of arithmetic and reading. Kirby and Woodhouse focus on the later skills of learning from text, and how these might be assessed. The following articles, by Nagy, Winne et al., Rogers and Bateson, and Randhawa, report a series of empirical studies that address the measurement of achievement in cognitive frameworks. The final articles, by Bateson and Bachor et al., revisit the issue of validity and return the discussion to the implications for classroom practice.

Acknowledgments

The conference on which this special issue is based was supported by the Social Sciences and Humanities Research Council of Canada, the School of Graduate Studies and Research and the Faculty of Education of Queen's University, whose help is gratefully acknowledged. We would like to thank Lyn Shulha for her help in organizing the conference, and Gay Bisanz, Phil Nagy, Randy Randhawa, and Phil Winne for their helpful comments on this introduction.

References

Anderson, J.A. (Ed.). (1990). Assessment in the classroom. Special Issue: *Alberta Journal of Educational Research, 36.*

Baker, E.L., O'Neil, H.F. Jr., & Linn, R.L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48,* 1210-1218.

Bateson, D.J. (1992). *Classroom testing in Canada.* Vancouver: Centre for Applied Studies in Evaluation (CASE), UBC.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO Taxonomy*. New York: Academic Press.

Carroll, J.B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L.B. Resnick (Ed.), *The nature of intelligence* (pp. 27-56). Hillsdale, NJ: Erlbaum.

Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Das, J.P., Kirby, J.R. & Jarman, R.F. (1975). Simultaneous and successive synthesis: An alternative model for cognitive abilities. *Psychological Bulletin, 82*, 87-103.

Ferguson, G.A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8*, 95-112.

Sternberg, R.J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 703-713.

Wittrock, M.C., & Baker, E.L. (1991). (Eds.). *Testing and cognition*. Englewood Cliffs, NJ: Prentice-Hall.

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 109-126*

Thomas Maguire
University of Alberta

John Hattie
University of Western Australia

and

Brian Haig
University of Canterbury

Construct Validity and Achievement Assessment

The psychometric notion of construct validity has evolved greatly since its introduction by Cronbach and Meehl (1955). Its current status is to be found in a carefully formulated article by Messick (1989) in which he embeds construct validity in a realist conception of social science. In this article the authors examine several facets of Messick's approach. They take issue with the emphasis placed on consequential validity as a component of construct validity, the inadequate treatment of the link between tests and scores, and the focus on construct validity in large-scale standardized test development. The authors show the profitability of separating the construct-indicator link from the indicator-score link and suggest that although the latter is not always necessary, its treatment in the construct validity literature has generally been inadequate.

Depuis les années 50, lorsque Cronback et Meehl (1955) ont introduit la notion psychométrique de la validité de la synthèse d'éléments constituant une construction intellectuelle, celle-ci a évolué grandement. Cette notion est présentée de nos jours de façon méticuleuse dans l'étude de Messick (1989) dans laquelle il présente cette notion dans le contexte d'un concept réaliste des sciences sociales. Dans cette étude, les auteurs examinent les diverses approches de Messick. Ils questionnent l'emphase que l'on met sur la façon dont la validité conséquentielle ou indirecte peut être une composante de la synthèse d'éléments constituant une construction intellectuelle. Ils questionnent également le traitement inadéquat des liens entre le testing et les scores. Ils interrogent le rôle que joue la validité de la synthèse constituant une construction intellectuelle dans le développement de tests standardisés sur une grande échelle. Les auteurs démontrent à quel point il est profitable de séparer les liens qui déterminent les indicateurs d'une construction intellectuelle des liens indicateurs des scores. Ils suggèrent que l'usage des indicateurs des scores, même s'il n'est pas toujours nécessaire, a tout de même été mal vu dans les écrits qui traitent la validité de la synthèse d'éléments constituant une construction intellectuelle.

Thomas Maguire is a professor of educational psychology. His research interests are in educational measurement and applied statistics.
John Hattie is a professor of education. His research interests are in educational measurement, self-esteem, and multivariate analysis.
Brian Haig is a senior lecturer in psychology. His interests are in philosophy of social science and in the study of research methods.

Messick's (1989) chapter in Linn's *Educational Measurement* must be taken as the most extensive explication of current thinking on construct validity. It aims at providing a "unified though faceted" approach to the notion of validity grounded in ideas about scientific methodology but located in the reality of the political context. The purpose of this article is to examine Messick's notion of validity as it applies to the assessment of educational achievement. Woven through our article are five related points of contention that we have with Messick's chapter as it applies to the construction, interpretation, and use of achievement tests. First, we believe that in spite of the lengthy description of his preferred realist philosophy of construct validity, Messick uses the approach inconsistently. Second, we contend that widening the notion of validity to encompass the notion of consequential validity is inappropriate. Third, Messick's chapter is unsatisfactory with respect to its emphasis on "scores and measurements as opposed to tests or instruments because the properties that signify adequate assessment are properties of scores not tests." Fourth, treatment of structural validity is inadequate. Fifth, the role of the measurement "establishment" is overemphasized at the cost of public accessibility.

The assessment of educational achievement and in particular the assessment of cognition is essentially about describing the unobservable. The focus of such assessment is on internal knowledge states, mental structures, and reasoning processes, and the branch of psychometrics that informs this work is construct validity. We make no apologies for writing this article as a commentary on Messick's (1989) chapter because we believe the chapter is a significant piece and will be a major reference for evaluating the outcomes of research in assessment of cognition.

The article is organized into five sections. In the first, features of Messick's chapter that pertain to our five basic assertions are presented. The second section examines the notion of consequential validity and takes the position that the presuppositions of consequential validity are too important to be treated just as a part of construct validity. A brief discussion of the substantive and external components of validity are presented in the third section, "From Construct to Indicator and Back." Following this, the section "From Indicator to Score" raises issues concerning the underemphasized, overlooked, and misunderstood aspects of structural validity. The concluding section attempts to draw the threads together.

Some Features of Messick's View of Construct Validity

It is fitting to begin with Messick's (1989) opening paragraph because it captures the essentials of both the connotative and denotative aspects of validity for him:

> Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. As will be delineated shortly, the term test score is used generically here in its broadest sense to mean any observed consistency, not just on tests as ordinarily conceived, but on any means of observing or documenting consistent behaviors or attributes. Broadly speaking, then, validity is an inductive summary of both existing evidence for and the potential consequences of score interpretation and use. Hence what is to be validated is not the test or observation device as such

> but the inferences derived from test scores or other indicators—inferences about score meaning or interpretation and about the implications for action that the interpretation entails. (p. 13).

For this article it is important to notice two central features in this quotation: test scores and potential consequences. The main issues of contention derive from the prominence that Messick assigns to each. Although "score" is used in its "most general sense of any coding or summarization of observed consistencies on a test questionnaire, observation procedure or other assessment device" (p. 14), it is important to remind ourselves that the ordinary meaning of the term and its most common application is as a special kind of *inductive summary*, a numerical summary or test score that results from a measurement. The title of the volume, *Educational Measurement*, tells us that we cannot ignore the importance of score in this conventional sense. What readers of previous versions of validity chapters in *Educational Measurement* will find surprising is the importance that Messick assigns to "the functional worth of scores in terms of social consequences of their use" (p. 13).

The unifying structure for Messick's chapter is the two-dimensional Facets of Validity Table in which the "Functions or outcomes of testing" (split into test interpretation and test use), are crossed with the "Sources of justification" (divided into evidential bases and consequential bases) to produce a fourfold table: (a) the evidential basis of test interpretation—construct validity, (b) the evidential basis of test use—construct validity incorporating relevance and utility, (c) the consequential basis of test interpretation—construct validity and value implications, and (d) the consequential basis of test use—construct validity, relevance, utility, and the value and social consequences. Messick talks about the cells as being a kind of progressive matrix moving from the classical Cronbach and Meehl (1955), and Loevinger (1957) approach to validity toward a conception that includes social implications. Although both evidential and consequential aspects of test interpretation precede their appearance as test use in Messick's presentation, we (like Shepard, 1993) prefer the sequential portrayal because it makes it easier to place a boundary between construct validity and what we believe is more properly a matter for professional ethics and social debate.

Test validity has its roots in the positivist-operationalist tradition, where intelligence and personality attributes are characterized in terms of testing operations. In a lengthy section, Messick shows how construct validity, the essence of all validity, is best embedded in a constructive-realist philosophy of science. According to him, "Realists view theoretical terms as conjectures about existing, though presently unobserved, attributes of the world" (1989, p. 26). A construct is the name that we give to the underlying mechanism that explains individual and social behavior, and so for the realist, test and non-test consistencies are explained by reference to a theoretical system of constructs. Construct validation aims at describing, in terms of the surrounding theory, the links between performance on assessment instruments, the nature and construction of the instruments, and the underlying structures and processes that are elicited by the instruments. Although it is possible to talk about validity without reference to a realist position, most current writing in the area is based on that perspective.

The current interest in relating research in cognitive psychology to educational assessment favors a realist position. If we adopt the stance that educational tests should be consistent with the goals and processes of education, if these goals and processes refer to mental and social entities called achievement, and if achievement is presumed to exist and serve an explanatory function (neither a hazy conversational metaphor like "being well dressed," nor a simple operationalist indicator like "number of on-task behaviors") that is in some way measurable, then we are in the realist camp, and the nature of achievement as a construct and its validation warrant our attention.

As it should, Messick's evidential basis of test interpretation borrows much from Cronbach and Meehl (1955) and the subsequent 40 years of construct validity literature. The three components of validity: substantive, structural, and external (Loevinger, 1957); the notion of construct validation as a continuing research process (Cronbach, 1971); the multitrait, multimethod approach to convergent and discriminant evidence (Campbell & Fiske, 1959); and the importance of "domain" (Lennon, 1956) are all to be found in this important section. To the methods of validation listed by Cronbach (1971) Messick adds discourse analyses of various sorts, confirmatory factor analysis, mathematical modeling, and even the analysis of eye movements. But the bulk of the discussion seems to ignore, or at least gloss over, some of the critical aspects: test items seem to appear by magic and are treated as givens; their combination into scores ignores the representational assumptions that would be required by a genuine realist approach; and the emphasis on modeling and components of test variance encourages our attention to drift from achievement about individuals to achievement as an abstraction about groups.

Although the section on the evidential basis of test use covering curricular validity, predictive validity, and utility is a friendly reminder of the psychometrics of Taylor and Russell (1939), Cureton (1951), Cronbach and Gleser (1965), and Ghiselli (1966), the sections on the consequential basis of test interpretation and the consequential basis of test use are a strong plea from Messick to accommodate the various social critics of educational testing by bringing the value orientations, ideologies, social consequences, side effects, and the ethics of testing into the construct validity arena. This is a major point of Messick's work and it is examined in the next section.

Consequential Validity

Messick (1989) claims that consequential validity involves the appraisal of both potential and actual social consequences of test score interpretation and use. It relates to the appraisal of the value implications of the construct label and of the theory underlying the test interpretation, as well as an appraisal of the ideologies in which the theory is embedded. In effect he argues that both the intended consequences and the side effects as seen from multiple perspectives should serve to govern our activities with respect to development, interpretation and use. It is our contention that the priority Messick assigns to consequential validity derives from a particular and narrow political social context, is inconsistent with the realist perspective he describes and endorses, and inappropriately locates certain aspects of professional responsibility with the test developer rather than the test user.

Although following Messick's advice might reduce liability for commercial test developers in a litigious society, such efforts to avoid suits would probably divert energy away from the developers' primary responsibility to collect and report evidence on how the test and its scoring system relate to the underlying construct. In addition, although passing acknowledgment is paid to the test user's responsibility ("a heavy ethical burden,") Messick does not spell this out at all. We believe that Messick's argument for consequential validity as central to validity fails for several reasons.

1. *There is an overemphasis on the role of the developer of large-scale standardized tests.*

If Linn (1989) and his associates had organized their book from the perspective of the test taker rather than developer, it would have been obvious that the CTBS, NAEP, TIMMS, and all the other externally developed achievement measures occupy a relatively small place in the testing lives of students. Teacher-built achievement instruments account for much more of this time, and one would expect the chapter on test validity in one of the significant source books on educational measurement to acknowledge this. The issues Messick raises about consequential validity of standardized tests when applied at the classroom level are really matters for the teaching profession to consider. From the point of view of individual teachers the consequences of their own tests for instruction and student learning are much more important.

Even if we concentrate on standardized instruments, Messick ignores how achievement tests come to be used. In many jurisdictions, national, provincial-state, or local tests are mandated, and they are constructed for specific uses such as accountability. It is true that test developers are often an external agency, but the constraints provided by the test users (i.e., the people or institutions that commission the tests) often set the ideological agenda for test development. So, for example, much of the political support (and funding) for the various international testing programs is to be found in the belief that countries must be competitive in a global economy. Governments think that by knowing how our children compare with the children of other industrial countries in relation to mathematics and science achievement, they can use the information to improve the relative economic position. This forms part of the ideological context that governs the use of scores. The test constructors take great pains in producing instruments that are culturally neutral, useful, and efficient. The values that drive the interpretation do need to be critically examined, but little of this examination will be directed at the activities of the test developers.

2. *There is an emphasis on large-scale, systematic and planned assessments of consequential validity.*

Messick's list of strategies for examining consequential validity makes sense if testing is viewed as an industry like the pharmaceutical industry, but as argued in 1 the bulk of student testing experiences are not of this sort. We believe that the role of test developer (commercial, public, or classroom) is to make the evidential basis for interpretation clear and not to make a list of approved and disapproved uses.

The implicit value assumptions and social consequences of testing are better examined through processes such as those raised by Kane (1992) who presented an argument-based approach to validity that is positioned to deal with how tests should be used and interpreted. Kane suggests that validation is argument that focuses on those parts of the interpretation of a construct that are most doubtful or problematic. He distinguishes two stages. The formative stage involves the clarification or explicit definition of the interpretive argument and the development of a preliminary case for its plausibility, that is, to "develop a plausible conjecture." The summative stage occurs when the interpretive argument is subjected to serious empirical challenges. The degree of confidence in the validity that is warranted depends on how well the interpretive argument survives the various challenges.

To illustrate Kane's (1992) ideas, consider a hypothetical example (although one that resembles an argument current in one provincial department of education). Suppose it is argued that a provincial achievement test in history is presented from a male perspective, such that it has an ideological gender bias. Sharpening this argument leads to two interrelated notions: that boys organize their thoughts on historical matters in a different way from girls, and that the test has been constructed to favor the male "approach." The difference in organization could be due to inherent cognitive differences, differential response to instruction, differences in motivation, differential responses to the test items, and so forth. Examination of the construction process (i.e., the activities engaged in by the test development committee) and statistical analyses of certain items suggest that the criticism is plausible. Using a think-aloud procedure to collect the reasoning processes with carefully selected boys and girls might subject the charge of gender bias to a genuine empirical challenge, and provide some insight into whether curriculum, instruction, and assessment should change, and how that change might be accomplished.

This procedure may not seem to be a reasonable part of the test validation process, because it depends on the test's use in a particular educational-social context to sensitize individuals to the potential problem. Yet once noted, the argument and its evidence could easily change testing practice. Underlying our position is the view that test interpretation is always a matter for debate. The goal of the test development enterprise is to make that debate available to developers, users (consumers), and even test takers where possible, by providing a clear description of the construct and explaining the evidence that links the scores to the construct.

3. *The emphasis on consequential validity favors test score use rather than test development.*

Early in his chapter Messick (1989) notes the importance of scores as "opposed to tests or instruments." Older conceptions of validity spoke of locating tests and constructs in law-like networks (nomological nets). As Shepard (1993) puts it,

> Old conceptions of validity were analogous to truth in labelling standards. A more apt metaphor for current validity requirements is the Federal Drug Administration's standards for research on a new drug. The scientist is responsible for evaluating both the theoretically anticipated effects and the side effects of

> any product before it is rendered "safe and effective" and released for wide-scale use. (p. 426)

This move from test to test use is too great. Even if we accepted consequential validity as an important component of validity, the process of test development would tell us much about the ideological bases of the test and its subsequent interpretation. But Messick's claim that validity is an "integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (1989, p. 13) tells us that what we do with the construct in a utilitarian sense is more important than what it is. In effect, Messick forsakes operationalism for realism only to succumb to a narrow pragmatic instrumentalism. From a realist point of view, a construct denotes an explanatory mechanism, and it may or may not have instrumental application at a particular time. However, our success in understanding the relevant educational phenomena depends in a large part on our effective development of the construct and its associated theory. If there is to be a good test, then an effective test-construct relationship is imperative. For this we must have good constructs. In the present context, achievement as a good construct demands appropriate theory linking knowledge, structures, reasoning processes, and so forth. Tests must derive from such theory, and their use must be consistent with it. However, as we shall see, there are many pitfalls to be negotiated before we get to test scores and their use.

4. *Much of the stress on consequential validity seems to be a reaction to the increased litigation surrounding tests and testing practice.*

Linn (1989) devotes about one quarter of his introduction to *Educational Measurement* to a discussion of the legal aspects of educational measurement, and makes a persuasive case that these concerns are vital to the field of educational measurement. Messick acknowledges this briefly toward the end of his chapter, and although he does not specifically locate his consequential arguments in this climate, it appears to have influenced his decision to assign consequential validity a role of major importance. But this provincialist lapse detracts from the realist position espoused by Messick at the beginning of his chapter. The realist is primarily concerned with fashioning good explanatory theories, and a concern with consequences will relate to that goal, but the greater emphasis will be on coherence, explanatory richness, and so forth.

We would not deny that the consequences of test use are important. Education in general and educational assessment in particular are and should be subject to social and legal scrutiny. But it is our contention that a concern with consequences should be moved out from the umbrella of construct validity and into the arena of informed social debate and formulated into ethical guidelines such as the *Principles of Fair Student Assessment Practices for Education in Canada* (1993). The *Principles* deal not only with items contained under Messick's consequential validity, but speak of follow-up and redress as well. They place ethical test use properly in the arena of professional responsibility and encourage an atmosphere of openness and questioning. If the debate on educational assessments is to be well informed, and if a social science of education is to become more than an oxymoron, we will need to devote our

psychometric attention to the evidential bases for test interpretation and test use. In turn the debate on consequences will rightly guide our subsequent test development activities just as it has with the "authentic" testing movement.

From Construct to Indicator and Back

In this section, we examine the link between constructs and their manifestations. We use the term *indicator* to refer to the responses or behavioral manifestations that might be elicited by an instrument, an observation, or any other testing situation. Indicators can be oral explanations, choice of alternatives in multiple-choice situations, laboratory performances, and so forth. We specifically exclude scores. The link between indicator and score is treated in the next section.

The validity of the link between indicator and construct has largely been studied by examining test performances and arguing back to their validity as signs or samples of the underlying achievement construct. Much significant work has been done in this area, and Messick (1989) acknowledges the importance of cognitive psychologists, the factor analysts, and the mathematical modelers. We are inclined to see the contributions of the first group as being more significant than the others because, as Manicas (1987) notes, causes are not likely to be additive, as the causal modelers commonly suppose. Models like linear factor analysis, or Embretson's (1983) model for response processes assume structural relations that mostly do not exist. Also, we must constantly be mindful that the hypothetico-deductive approach to construct testing cannot carry the entire inferential burden because a single data set can be consistent with two or more contradictory models. Moreover, the emphasis on validation evidence flowing from indicator to construct may subtly encourage an operationalist view of constructs wherein the indicator actually *becomes* the construct. Realists keep the construct/indicator distinction firmly in mind. The construct is the underlying mechanism that explains not only the behavior on a specific indicator, but other, non-test manifestations as well. We argue that the justification process needs to flow in both directions if a coherent realist view of achievement is to be sustained.

There is much to be said about the validity of the link between construct and indicator, but we discuss only two points. The first is the problem of legitimizing the developmental procedures involved in going from construct to indicator. The second concerns the constraints of test format on validity.

From Construct to Indicator

Traditionally, achievement test developers have been advised to adopt a table of specifications strategy that cross-tabulates Bloom, Englehart, Furst, Hill, and Krathwohl (1956) taxonomic categories with content areas. Items are then constructed and placed in categories in proportion to a desired emphasis. This encourages efforts toward content coverage on the part of the test developer, but it fails to take into account some fairly fundamental problems. For example, the Bloom model takes a static view of content, treating it as independent of cognition. As a result, issues such as context salience, position of specific pieces of content in the instructional flow, their position in the overall content structure, and the essential interactive nature of cognition and content tend to be ignored.

More prescriptive approaches to item design that begin with domain specifications and item construction strategies (e.g., Bormuth, 1970; Guttman, 1969; Hively, Patterson, & Page, 1968; Osburn, 1968) are denigrated by Messick (1989) because they "merely tell us at most that the test measures what the universe measures" (p. 40). Yet, on the face of it, it would seem that item generation procedures should speak to the validity of the link between construct and indicator.

Haertel (1985) was one of the first people to examine achievement as a construct. He argues that educational outcomes should be thought of as constructs:

> They are artificial, designed deliberately to equip students with a repertoire of appropriate responses to the complex settings and symbols of our culture. They are attributes defined primarily in terms of their behavioral manifestations and only secondarily in terms of cognitive processes or memory mechanisms. (p. 28)

Despite appearances this does not satisfy a realist orientation. Haertel's second sentence gives the game away. From a realist perspective an educational outcome would not be seen as an explanatory construct. Achievement as a construct would be described *primarily* in terms of internal processes and mechanisms, and these would be seen to give rise to certain behavioral manifestations. Educational *outcomes* are obviously important, and they are often described in behavioral terms. But the cardinal goals of education are tied to the development of internal achievement constructs. Snow (1989), for example, lists five broad achievement constructs: deep understanding, efficient intuitive use, multiple flexible strategies, adaptive action control, and achievement motivation.

For the realist the main problem for assessment is to develop and link indicators to the networks of such constructs. One approach to achievement assessment that exemplifies something of the realist view is Biggs and Collis's (1982) SOLO taxonomy. SOLO, the Structure Of Learning Outcomes, is a system for categorizing student responses according to their degree of elaboration, organization, consistency, capacity, generalization, and abstraction. The levels of the taxonomy can be thought of as stages of learning that an individual passes through in mastering new concepts. The levels of the taxonomy used to evaluate the quality of learning are described in terms of the amount of memory capacity required to make the response, the kind of relating operations used to produce the response, and the degree of consistency and the quality of closure displayed in the response. The application of the SOLO hierarchy is centered on declarative and procedural knowledge, but the attainment of higher levels presumes appropriate motivational orientations, self-regulatory functions, and learning strategies. The taxonomy is one promising attempt to represent the relationship between knowledge and its organization as the student is prepared to present it under test conditions. Moreover, it is a framework that can be used from classroom tests to national assessments.

The Constraints of Test Format

In the previous section, we talk about the construct-indicator link. In the case of achievement measurement, the indicator is usually a response to a task or an item. Like many others, we believe that test format can influence a test taker's

approach. The format influence may be a source of invalidity, but it is not self-evident. If we place formats along a continuum according to the degree to which the test taker's response is apparently constrained, multiple-choice questions would fall at one end and performance tasks would fall toward the other. But constraint on response may be inversely related to an item's "representational validity" (how well it represents the achievement construct). Complex performance-based assessments place great emphasis on problem solving, critical thinking in context, and other "higher order" thinking skills. As Linn, Baker, and Dunbar (1991) point out, this may not guarantee validity but it is a good start. Frederiksen and Collins (1989) argue that there are validity advantages to tests that evaluate cognitive skills "directly." Such tests obtain "systemic" validity because instruction that improves test scores will also improve performance on the extended task.

But more constrained formats like multiple-choice forms do not make the link simple. The argument that multiple-choice items only assess recognition simply cannot be credited by anyone who has made the effort to find out what processes students use when responding. To show this clearly, Skakun (1993, personal communication) asked senior medical students to think aloud as they responded to several multiple-choice questions in internal medicine. Three example answers to one of the questions are shown in Figure 1. Reading through the responses it is easy to see that a single question has tapped into an incredible variation in array of internal structures. We believe this illustrates the main problem with the construct validity of the multiple-choice format. Far from being a direct method for assessing recognition, this format is well capable of eliciting internal responses that are extremely complex. The problem is that these responses are not entirely foreseeable. The positive features of multiple-choice items are well known. But the one that is most supportable may be the efficiency of content coverage. If this feature is highly valued in a specific context, Willson's (1989) advice is to link each of the distractors to the process and knowledge requirements of the achievement construct.

As Snow and Lohman (1989) tell us, much important work is being done on the validity of the construct-indicator link. Our caution is to keep the bidirectionality of evidence in mind when thinking about this link, and to be aware of how the form of the indicator might influence the way it links with the underlying construct.

From Indicator to Score

One of Loevinger's (1957) three components of validity is structural validity. She describes it as the

> extent to which structural relations between test items parallel the structural relations of other manifestations of the trait being measured.... [it] includes both the fidelity of the structural model to the structural characteristics of non-test manifestations of the trait and the degree of inter-item structure. (p. 98)

This means that internal characterization of the construct should be represented in the indicator. For example, if we accept the metaphorical view of achievement as milestones along an educational highway, then the achievement indicators should form a scale. But if we see achievement as the emergence of stages through which an individual passes as he or she masters,

Item

A 56-year-old man presents to his doctor with a month history of intermittent facial pain. On examination he is found to have a diminished corneal reflex and slight hearing defect on the right. The diagnosis is:

1. Right cerebral tumor
2. Trigeminal neuralgia
3. Otitis media
4. *Acoustic neuroma
5. Multiple sclerosis

Response from Examinee A:

56-year-old man, they want a diagnosis, comes to doctor one month, history of right facial pain. I am thinking neurology. Diminished corneal reflex, slight defect in hearing. So we have facial pain on the right and pain is carried by the 5th cranial nerve, diminished corneal reflex again the 5th cranial nerve and he has hearing defect. Hearing defect, it could be sensory or conduction, so it could be related or not. If it is related it would indicate 5th cranial ... I mean 8th cranial nerve so there would be more that I would want to do on examination. I would want to find out how the 7th nerve was—whether he had any sort of ocular—extraocular defects—motor component but I guess if all they found on examination was no extraocular involvement just cranial nerve 5, I would go to the choices right now. Right facial pain—Right cerebral tumor, trigeminal neuralgia, otitis media, acoustic neuroma, multiple sclerosis. His age is getting up there for MS, acoustic neuroma has nothing to do with this corneal reflex unless it's impinging on the 5th cranial nerve ... one month history of intermittent facial pain, I'm going back up because obviously I don't know the answer right off, so I'm looking at the stem for clues I missed—like intermittent and diminished corneal reflex and slight defect on hearing so these are all, they are not cut and dried symptoms, they come and go or just decrease so it could be conduction problems and they are separated. I would want to know if he has had a past history of anything like this. Trigeminal neuralgia ... well, I know tic douleureux can present with facial pain but I haven't heard of hearing defect with it, but it's on the differential. So right now the cerebral tumor ... no, because it's in the head, up top, supratentorial and I'm looking a a cranial nerve defect. Trigeminal neuralgia—that's possible. It only usually ... it's usually the mandibular branch of the 5th cranial nerve. Otitis media, no there wouldn't be a problem with the corneal reflex. Acoustic neuroma, it's possible, but it shouldn't affect the 5th cranial nerve. Multiple sclerosis, that's on my differential so I'm stuck between trigeminal neuralgia and multiple sclerosis. I'll pick multiple sclerosis because of the distance between the nerves affected. I don't see the 5th cranial nerve coming out of the pons traversing the skull going through the foramina—I don't see it associated.

Response from Examinee B

A 56-year-old man, so he's middle-aged. He has a one month history of intermittent facial pain, to me ... right away I think of a trigeminal ... some kind of neuralgia. On examination he's found to have a diminished corneal reflex, so to me that means his 5th cranial nerve is involved. And slight hearing defect on the right and to me hearing defect in an 8th cranial nerve abnormality, and facial pain is also 5th nerve. So his 5th nerve and his 8th nerve are involved. So I'm looking in the answers for something that would cause that.

I'm looking at the answers and number 5, I don't think it's multiple sclerosis, its not typical presentation of multiple sclerosis. Otitis media, I'm looking at number 3, I don't think that's likely, the symptoms don't really correspond with that. Trigeminal neuralgia, I think it could be except for the fact that he's got hearing involved. So, and right cerebral tumor, it seems like it's cranial nerves that are involved and not the cerebrum so I would exclude that. So I would say number 4, acoustic neuroma, but I don't know why his facial nerve isn't affected, nerve number 7, but I would say either number 2 or number 4 and I'm going to pick number 4 because number 2 involves only the 5th nerve. Then again, I can't decide because it looks like ... I'm having a hard time deciding between answer number 2 and answer number 4 cause it looks like it could be ... like it could be trigeminal neuralgia except that he's got a hearing defect but that could be a

Figure 1. Example item and think-aloud responses.

separate thing. So, but I think neuralgia only leads to pain and not really diminished corneal reflex, I'm not sure but I'm going to stick with number 4 still.

Response from Examinee C

A middle-aged man with right facial pain and on examination diminished corneal reflex and slight hearing defect ion the right. Um, anytime there is an asymmetrical ... here in the absence of ... acoustic neuroma which is down here in number 4.

Right cerebral tumor is not in the right area to be cerebellar. Trigeminal neuralgia, there should be pain—there is some pain but it wouldn't explain the hearing defect. Multiple sclerosis is possible but unlikely in a gentleman of this age. Otitis media—unlikely. Acoustic neuroma in this gentleman until proven otherwise—number 4.

Response from Examinee D

A 56-year-old man presents ... slight hearing defect on the right. So it doesn't say where the right facial pain is caused if it's trigeminal neuralgia it would be greater in V2 than V3 then V1. He has diminished corneal reflex and a slight hearing defect on the right—hearing defect so 5 and cranial nerve 8 are being affected. Diagnosis of right cerebral tumor, no I don't think you can say that just from the history, number 1 is out, number 2 is possible, number 3, otitis media, well it doesn't explain the neurological findings, number 4 acoustic neuroma, boy oh boy, that's a possibility, number 5 MS he's not in the right age group. So between 2 and number 5—sorry 2 and 4, and a slight hearing defect on the right, I think with acoustic neuroma that hearing defect would probably come on before the corneal reflex, I think it would be pretty marked if you had a diminished corneal reflex so I'll say number 2.

Figure 1 (continued).

organizes, elaborates, and abstracts declarative and procedural knowledge, then the indicators should form a some kind of linked hierarchy. That Messick's (1989) comments on the indicator-score relationship are essentially those of Loevinger gives the misleading impression that this aspect of construct validity is not problematic. This is unfortunate because it is the central issue in the *measurement* of achievement (i.e., the link between indicator and score). Much effort has been expended on developing a technology of educational measurement, as witnessed by the large portion of journal space devoted to item response theory and related topics, but surprisingly little attempt has been made to integrate that technology with the overriding concerns of construct validity. Before expanding on this point, it is useful to step back and raise some questions about basic measurement theory.

Let us take as an initial characterization of measurement, the mapping of objects or phenomena onto a numerical continuum according to the variations in their characteristics. There is much we can do when we have described phenomena in numerical terms, and the rewards are so compelling that we often neglect to ask the important prior question, Is this phenomenon measurable? To answer the question, we need to know much about the object or phenomena, and we need to have a position on the nature of the mapping rules. As Michell (1990) points out, there are two conceptions of measurement: operationalism and representationalism. According to operationalism, concepts are indistinguishable from their corresponding sets of operations; measurement is simply a set of procedures that generates a number. From a realist point of view, this is not an acceptable position.

According to Michell (1990), under the representational theory of measurement, "Numbers are assigned to empirical entities in such a way that certain relations between the numbers assigned represent empirical relations between

the entities. Representationalists make a distinction between empirical entities and numerical ones" (p. 29). As applied to achievement, this means that there is a distinction between a test score and its construct. Berka (1983) makes the obvious distinction between physical measurement and extraphysical measurement as is found in the social and behavioral sciences. Measurement of constructs like achievement clearly fall in the latter category. Berka points out that extraphysical measurement is methodologically tied up with classification (dividing objects into sets according to some qualitative aspects) and with observation, and although it proceeds in a different manner from the usual experimental basis of physical measure, there are three common components: the object of measurement (in our case achievement), the results of the measurement (the value on the scale), and the instrument or the mediating empirical operations (for us the achievement test and its scoring system). If we are to work from Berka, we might be advised to keep the prior notion of classification in mind as we move toward valid "mediating empirical operations" (i.e., tests!).

Scoring models from classical test theory and from item response theory (IRT) provide two popular but distinct structural models of achievement. Under classical test theory, building tests from discrete items and adding the item scores together is consistent with a metaphorical view of achievement as a long coin tube, where each piece of achievement is a coin. A student's score is given by the number of achievement coins that he or she possesses. The presence of this or that coin and their order in the tube are not important; it is the sum of the item scores that indicates the amount of achievement. The IRT version of achievement is like the highway described earlier. Items are like milestones along the road. The achievement score tells us at which position the student is to be found. Under the Rasch scoring rules of IRT, these summative and positional approaches yield the same results, but that is a consequence of the mathematical assumptions of the procedure and not an ontological assumption about achievement itself.

To appreciate the distinction, consider the example of the Alberta Infant Motor Development Scale (Piper, Pinnell, Darrah, Maguire, & Byrne, 1992). The scale is designed to assess motor development in infants from birth to independent walking. The items form an almost perfect Guttman scale. Interestingly enough, however, if an infant is located at a particular point on the scale he or she may no longer be able to successfully perform items from a previous point. Those items (i.e., earlier items) may have been fundamental to healthy development, but they are lost or incorporated into higher level skills. Total scores on items would yield different positions for infants than milestone scores. We know that the latter has greater construct validity because it conforms to the most coherent theories of infant motor emergence.

Examples of achievement constructs that fit either the classical summative scoring model or the item response continuum model are less common than the advocates of either approach would have us believe. Without surrendering to the operationalist camp, we agree that observables and the empirical models on which they are based must influence the notion of the construct, but in the case of most testing programs this has gone too far. The scoring model en-

courages a view of achievement that is simply inconsistent with most of our current understanding of achievement.

Portrayals of the achievement construct such as that given by Snow (1989) prompts us to question whether achievement is measurable at all. Remembering Berka, we should think in terms of classification of achievement types. Perhaps the criterion-reference people have been correct all along with their decision approach to achievement assessment. Although it did not take long to convert the simple ideas of Glaser (1963) into the mathematics of cut-off scores, loss functions, and domain estimates (Hambleton, Swaminathan, Algina, & Coulson, 1978), the essential notion of competence is still of use. The configuration, pattern, or position that a person possesses in relation to the achievement construct of interest is characterized via the indicators. Our initial role as assessors is to sort these portrayals into groups that are similar. Judgments can then be made on whether a particular group possesses "mortal deficiencies." This is not yet measurement, but neither does it distort the construct representation due to an inappropriate and arbitrary scoring system.

If we want a measurement, and we suspect that this should be far less often than current practice suggests, we need to think about defining a single continuum. Achievement as Snow (1989) and others construe it is complex and multifaceted, so that indicators of achievements might be ordered differently in relation to different facets. For example, if we wanted to look at achievement with respect to mathematical problem solving and we administered several open-ended mathematical problems, then we might consider facets like efficiency, creativity, transferability, conformity to taught patterns, and so forth as being important, and the performances on these facets would have different positions relative to each other. The scaling solutions would require assumptions to be made, but care could be taken to ensure that they retained a reasonably plausible representation of the facet of the construct.

If decisions require a single ordering of people, then the nature of the decision should guide the scaling process. If the decision requires only categories, then the combining of indicators need not produce a score. Take a complex example to illustrate. Each year the Medical Council of Canada administers a licensing examination to about 1,700 graduating medical students from across the country. In addition, approximately 700 graduates from outside Canada take the exam. Three quarters of the examination is made up of multiple-choice questions assessing competence in six disciplines (medicine, surgery, pediatrics, obstetrics and gynecology, psychiatry, and preventive medicine and community health); the remainder of the examination consists of patient management problems. Until recently, for candidates who wrote the exam, the number correct on the multiple-choice section was combined with appropriate weighting to the patient management section, and the candidates were ordered. The 4th percentile of the Canadian student performance was used as the licensing criterion.

If we think of the multiple-choice portion of the examination as a large achievement test, with each item testing some complex aspect of medical training, we might ask ourselves if there is a more valid way of arriving at the decision to license or not. Notice that there is no inherent reason for recasting the problem as a cutting score problem.

The multiple-choice items are selected from a bank or are created by the six disciplinary test committees. Each committee is composed of six members who are a careful blend of discipline specialists, specialists in family medicine, faculty of medical schools, and general practitioners. Each of these committees participated in a modification of the Nedelsky standard setting procedure using a kind of Delphi approach (Maguire, Skakun, & Harley, 1992). The goal of the process was to reach consensus judgment on each alternative to each item. The basis of the judgment was that a graduating medical student who is about to enter supervised clinical practice should know whether an alternative is plausible or not. Without apology, the arguments and judgments were based on the experience of the panels, the importance of the concept (e.g., is it fundamental, is it dangerous to be wrong about it?), and the knowledge that passing candidates would be allowed to take internships anywhere in the country.

As noted above, each of the items is capable of eliciting a complex sequence of thought and there are many ways of getting an answer correct. But if the thrust of the decision is to eliminate people with inappropriate conceptions, a scoring system could have been built up from patterns of responses on these items, so that, for example, it might turn out that the panels believe that any candidate who makes a particular pattern of errors should not be granted a license to practice. Several patterns of responses might be judged inadequate, and adjustments could be made to fit the uncertainties of the testing situation. If it turned out that a cutting score on the simple number correct yielded the same decisions (and it might), then much effort could be saved by doing that. But the justification would lie in the fact that it achieved the same results as a process that retains representational fidelity.

Essentially our thesis is that tests and scoring procedures are *mediating empirical operations* to make constructs into numbers. If the things that we do with those numbers (make decisions, conduct research, etc.) are to have any validity, then we must take great care in developing the tests and scoring procedures. It is not obvious that the common practice of administering large numbers of items and computing number correct, or weighted aggregates, retains the representational link between scores and constructs. Given the complexity of achievement constructs we think it will be often more useful to think in terms of ordered categories.

Concluding Thoughts on the Construct Validity of Achievement Tests

At the end of the day we must ask ourselves if the realist view of achievement is worth the effort. We think the answer is Yes. This is because of our belief that educational social science advances on the strength of well-developed theoretical constructs. The importance of achievement stems from it being more than an indicator; it essentially involves changes in the way students think and understand. Pursuing achievement indicators that reflect deeper structural changes, that is, that possess construct validity, is worthy of our attention.

We have tried to establish the importance of two links: construct-indicator, and indicator-score. Although they are blurred in most of the educational measurement literature, we think the distinction between them is useful in helping us to see the greater importance of the first. In considering these links we are led to the conclusion that many scoring systems in use have not been

validated in Loevinger's (1957) structural sense, and that it may not be necessary. However the urge to move to scores is ingrained in our thinking. The test validation procedures in common use assume a score. Factor analysis, the traditional workhorse of test validation, and structural modeling, its likely successor, seem like powerful procedures. But they have their limitations. They make metric assumptions that have consequences for our understanding of the nature of the relationships among the constructs they help us to investigate, but these assumptions are wantonly disregarded in practice. In addition, they tend to encourage sloppy talk about the nature of causal relationships.

The center of that sloppiness is to be found in the notion of variance. In the Messick (1989) chapter its use is pervasive. Shared variance, test-irrelevant variance, and so forth are central to the discussion of evidential aspects of validity. But to be pedantic for a moment, variance is a description of dispersion in a group of scores. Even at a superficial level, to talk about sharing variance, or apportioning variance, requires adherence to a particular kind of model (usually additive) that may well misleadingly depict how constructs relate to each other in an individual or a group of similar people.

We believe that the enlightened study of achievement and how it evolves will involve the intensive study of one or a few people at a time. Alton-Lee and Nuthall (1992) conducted a series of studies that provide an excellent illustration. They observed and recorded all the student teacher interactions, student-student interactions, and individual student vocalizations and sub-vocalizations that occurred during the teaching of an entire unit of instruction. They used four different units of instruction (ranging in time from 5 days to 31 days) and reported their results on 14 different students (3 or 4 per unit). By looking at pre-test, immediate posttest and 12-month posttest changes on achievement items (from 81 to 113 per unit), they classified outcomes as: items that were already known; items that were not learned; items that were learned and forgotten; and items that were learned and remembered. From their observations on two students, they produced a model that described student learning as the creation of specific knowledge constructs. The model defined the number of instances and types of relevant experience required for construct generation and the time intervals in which the experiences must occur. Using the model on other students they were able to predict with over 80% success student learning (or not) on individual test items.

Studies like those conducted by Alton-Lee and Nuthall (1992) tell us much about how classroom activity influences learning, but to improve our understanding of achievement we will need to become even more penetrating at the micro level. The strategies of investigation will be as Berka (1983) notes: logical, observational, and qualitative. The gains from this should be worth the effort.

Tests are often thought of as the tail that wags the educational dog. But this metaphor is too glib. The consequences of educational assessment both for individuals and for society are important. Testing works in support of education, and our position underlines the need for their principled use, but construct validation as it applies to achievement assessment should mainly be directed at promoting a real understanding of the nature of achievement constructs and how they might be represented in assessment. To focus on the

consequential use of tests is to encourage a view that educational stakeholders are relieved of their responsibility to scrutinize all facets of the educational enterprise in the context of emerging social change.

References

Alton-Lee, A., & Nuthall, G. (1992, April). *Students learning in classrooms: Curricular, instructional and sociocultural processes influencing student interaction with curriculum content*. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco.

Berka, K. (1983). *Measurement, its concepts, theories and problems*. Boston: Reidel.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning, The SOLO Taxonomy*. New York: Academic Press.

Bloom, B.S. (Ed.)., Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *A taxonomy of educational objectives: Handbook I: Cognitive Domain*. New York: Longmans, Green.

Bormuth, J.R. (1970). *On the theory of achievement test items*. Chicago: University of Chicago Press.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Cureton, E.E. (1951). Validity. In E.F. Lindquist (Ed.), *Educational measurement* (pp. 621-694). Washington, DC: American Council on Education.

Embretson, W.S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.

Frederiksen, J.R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Ghiselli, E.E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist, 18*, 519-521.

Guttman, L. (1969). (Ed.). Integration of test design and analysis. *Proceedings of the 1969 Invitational Conference on Testing Problems* (pp. 53-65). Princeton NJ: Educational Testing Service.

Haertel, E. (1985). Construct validity and criterion-referenced testing. *Review of Educational Research, 55*, 23-46.

Hambleton, R.K., Swaminathan, H., Algina, J., & Coulson, D.B. (1978). Criterion-referenced measurement: A review of technical issues and developments. *Review of Educational Research, 48*, 1-48.

Hively, W., Patterson, H.L., & Page, S.H. (1968). A "universe defined " system of arithmetic achievement tests. *Journal of Educational Measurement, 5*, 275-290.

Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.

Lennon, R.T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement, 16*, 294-304.

Linn, R.L. (1989). Current perspectives and future directions. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 1-10). New York: American Council on Education and Macmillan.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-23.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694 (Monograph Supp. 9).

Maguire, T.O., Skakun, E.N., & Harley, C. (1992). Setting standards for multiple-choice items in clinical reasoning. *Evaluation and the Health Professions, 15*, 434-452.

Manicas, P.T. (1987). *A history and philosophy of the social sciences*. Oxford: Blackwell.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.

Osburn, H.G. (1968). Item sampling for achievement testing. *Educational and Psychological Measurement, 28*, 95-104.

Piper, M.C., Pinnell, L.E., Darrah, J., Maguire, T.O., & Byrne, P.J. (1992). Construction and validation of the Alberta Infant Motor Scale. *Canadian Journal of Public Health, 83* (Supp. 2), s46-s50.

Principles of fair student assessment practices for education in Canada. (1993). Edmonton, AB: Joint Advisory Committee. Centre for Research in Applied Measurement and Evaluation, University of Alberta.

Shepard, L.A. (1993) Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of Research in Education, 19,* 405-450.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*(9), 8-14.

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.

Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology, 23,* 565-578.

Willson, V.L. (1989). Cognitive and developmental effects on item performance in intelligence and achievement tests for young children. *Journal of Educational Measurement, 26,* 103-119.

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 127-146*

Gay L. Bisanz

and

Jeffrey Bisanz
University of Alberta

Studying the Development of Academic Skills in the 1990s: Implications for Assessment

Psychological research on the development of academic skills has implications for educational assessment. We review research in two areas that typify the enterprise, arithmetic, and early reading, and we identify some of these implications. One conclusion is that productive interactions between assessment and cognitive development may well depend on the degree to which these specialities collaborate to produce individuals with the skills and insights that can arise from an interest and expertise in both domains.

La recherche psychologique sur le développement des habiletés académiques a une certaine portée sur l'évaluation en éducation. Nous passons en revue deux domaines de la recherche qui traitent de l'évaluation en arithmétique et en lecture hâtive et nous identifions quelques-unes de ces portées. Une conclusion établit que les interactions productives entre l'évaluation et le développement cognitif peuvent bien dépendre du niveau par lequel ces deux spécialités collaborent afin de développer des individus possédant des habiletés et une perspicacité qui proviennent d'un intérêt et d'une expertise dans les deux domaines.

Educational assessment in Canada and the United States is sustained by a variety of stakeholders—including politicians, administrators, teachers, parents, and students—with diverse and sometimes conflicting needs. Some stakeholders are concerned about systemic evaluation for purposes of public accountability. Their goal is to meet public standards for education while minimizing waste and expense. Others are opportunists, using test scores to sell homes in desirable neighborhoods or hawk the latest textbook series. Still others, like parents and teachers, have in mind particular children in their care. They value assessments that will help these children learn in school. Public debate among stakeholders about assessment often seems related to the unlikely but efficient premise that one type of assessment can be developed that will be suitable for all needs. We are skeptical about this premise. Rapid advances in new methods of assessment may well depend on the extent to which stakeholders come to recognize the diversity of needs and support the development of multiple methods of assessment to serve them.

Gay Bisanz is an associate professor of psychology and Director of the Centre for Research in Child Development. She studies cognitive development with special interests in early literacy skills, text comprehension, and science instruction.
Jeff Bisanz is an associate professor of psychology and a fellow of the Centre for Research in Child Development. He studies cognitive development and is engaged in research on the development of mathematical skills and on scientific reasoning.

Our goal is to review recent research on cognitive development with a focus on two areas of study, arithmetic and early reading, and to explore implications of this research for the development of new assessment techniques. Most of this work is focused on processes of knowledge acquisition, including the mental processes involved in cognition and motivation as well as the social and cultural processes that characterize the environments in which children learn. As we describe this research, it becomes clear that the contributions of research on cognitive development to educational assessment are likely to be greatest for stakeholders whose primary concern is the optimization of learning in individual children. These stakeholders include teachers, parents, and the students themselves.

Since the turn of the century and the work of G. Stanley Hall, developmental psychologists have shared an interest with many members of the general public in understanding and optimizing children's growth and development (Cairns, 1983). Widespread work related to the development and optimization of academic skill is relatively recent, however. A review of research on cognitive development in the 1960s would reveal a substantial number of articles on learning using traditional laboratory tasks, as well as a flurry of investigations focused on exploring and extending Piaget's theory. These lines of research were supplemented by the growth of a significant body of work on the development of memory, inspired in part by models of memory in adult cognition but with orientations and methods that reflected the special challenges of studying development rather than performance (e.g., Kail & Hagen, 1977). The use of information processing frameworks and methods was in its infancy, having been suggested by Simon in 1962. Despite growing interest in learning and cognitive development in children, research specifically designed to investigate the acquisition of academic skills was rare. For example, very few pages in Stevenson's (1972) capstone book on children's learning were devoted to research on academic skills.

Today neither traditional learning theory nor Piagetian theory provides a unifying framework for research on cognitive development. Information processing approaches and various contextual theories are important but not dominant. Research on cognitive development in specific domains (e.g., language, mathematical reasoning, naive theories in science) is clearly ascendant, but it is eclectic with respect to theoretical perspective. Despite the focus on development in specific domains, many developmental researchers have not abandoned Piaget's search for unifying principles in development. This agenda is represented, for example, by Kail's (1991) research on age-related changes in processing speed, Case's (1993) search for central conceptual structures, and Siegler's (1986) analyses of common developmental processes across different domains of cognitive skill. Renewed interest in Vygotsky's work and, more generally, on the role of social processes on cognitive development (Wertsch & Kanner, 1992) reflects a growing concern with the imbalance arising from the Piagetian metaphor of child-as-scientist. Establishment of the *International Journal of Behavioral Development* and the growth of the International Society for the Study of Behavioral Development reflect an increased awareness of the importance of culture in the study of cognitive development.

In this setting, developmental psychologists have begun (a) to characterize the acquisition of specific academic skills, such as reading, writing, problem solving, and reasoning in subject matter domains, and (b) to identify the effects of schooling as a set of experiences that shape cognition. The work on arithmetic and early reading typify the enterprise. We review research on arithmetic to illustrate the interplay between methods and theories used in research to assess cognitive skills (see also Royer, Cisero, & Carlo, 1993). Research on early reading is used to illustrate the perspectives afforded by new developmental theories and the factors that affect learning in children's environments. Finally, we examine a number of implications for assessment that arise from this research.

Research on the Development of Arithmetic Skills

Research on the development of mathematical cognition has mushroomed over the past two decades, largely because of methodological and theoretical advances that have enabled investigators to address important questions about the relations among general remembering and problem solving processes and knowledge of material in a specific domain. For purposes of illustration, we focus primarily on research designed to explore how children and adults solve single-digit arithmetic problems. Proficiency in simple arithmetic is only a component of mathematical cognition, but it is common, practical activity that is likely to be important for fluency and motivation in more complex forms of problem solving and reasoning.

Doing Arithmetic

Identifying procedures. The traditional approach to assessing proficiency in arithmetic is heavily oriented to the *products* of children's thought, as opposed to the *processes* that generate products or answers. Children or adults typically are presented with a set of arithmetic problems and the number of problems solved correctly is determined, often within some limited period of time. So, for example, a student who provides the set of answers given in Figure 1 would be judged to have solved 60% of the items correctly.

In contrast to traditional assessments of achievement, cognitive approaches are heavily oriented to identifying the mental processes that underlie performance. The distinction between product and process orientations is highlighted by a careful examination of the answers given to the problems in Figure 1. Notice that all the answers, correct and incorrect, reflect the operation of a consistent, underlying rule, according to which it is appropriate to subtract the smaller value from the larger value, regardless of which number is on top and

(a)	(b)	(c)	(d)	(e)
823	632	259	389	426
-332	-220	-127	-229	-349
511*	412	132	160	123*

Figure 1. Five subtraction problems. Incorrect answers are indicated with an asterisk.

which is below. Erroneous computational rules of this type are often referred to as *bugs* because, like bugs in computer programs, they are responsible for inappropriate output. A number of such subtraction bugs have been identified in children's arithmetic (Brown & Burton, 1978; Brown & VanLehn, 1982). Notice that the extent to which a bug produces erroneous output depends on the problem to which it is applied. Despite the bug, a child might obtain a high score if the problems on a test are similar to problems (b)-(d) in Figure 1. Even if performance is not perfect, a high score might serve to strengthen the bug, especially if the child views arithmetic as a maze of arbitrary rules that can never be negotiated with complete success.

The distinction between product and process is not unique to contemporary cognitive psychology (e.g., Werner, 1937), but cognitive and developmental psychologists have developed a variety of methods to identify the processes used on academic tasks (Kail & Bisanz, 1992). Analyses of response accuracy, similar to our example with subtraction, are instances of a method known as *rule assessment* (Siegler, 1976, 1981). Researchers also examine the solution time or *latency* to identify the procedures used by students. For example, young children often use counting-based procedures to solve simple arithmetic problems. According to a sum procedure, when a child is presented a numerical problem of the form a+b, he or she increments a mental counter a total of a+b times. Each such increment presumably takes a certain amount of time, and so solution latencies should vary as a function of the sum of a and b. According to an alternative hypothesis, the child sets the mental counter to the value of the larger number, a or b, and then increments the mental counter a number of times corresponding to the smaller of the two digits. Consequently, latencies should vary as a function of *min* (a,b), rather than *sum* (a,b). Hypotheses such as these can be tested by comparing latencies for different problems (Ashcraft, 1982).

Finally, analyses of behavioral observations and self-reports are also conducted to identify the procedures used to solve problems. While solving addition problems, for example, children often can be observed to count on their fingers and, when asked, can describe a counting procedure. Self-reports are particularly useful for obtaining insights about solution procedures when children use methods, such as retrieving answers from memory, that are not likely to have behavioral manifestations (e.g., Bisanz, Morrison, & Dunn, in press; Siegler, 1987).

Thus researchers have relied on several different kinds of dependent measures to identify how children solve arithmetic problems. Each of these methods has limits. For example, rule assessment methods can be ineffective if students do not use a single rule consistently (Kail & Bisanz, 1982), analyses of latencies can be qualified by patterns of error data (Pachella, 1974) or by variability in the procedures used (Siegler, 1987), and the validity of self-report data may be limited under conditions that are not as yet well specified (Cooney & Ladd, 1992; Russo, Johnson, & Stephens, 1989). Given these problems, the most compelling conclusions typically arise from studies in which the results from different dependent measures converge (e.g., Siegler & Shrager, 1984).

Three clear conclusions have arisen from this research. First, solution procedures change considerably as children become more skilled. For example,

younger, less skilled children tend to use counting-based procedures, such as the *sum* and *min* procedures, whereas older children and adults tend to rely primarily on direct retrieval of arithmetic facts from memory (e.g., Ashcraft, 1982; Ashcraft & Fierman, 1982). Second, individuals typically use a variety of different solution procedures to solve problems rather than being slavishly attached to a single procedure (Siegler, 1987; Siegler & Shrager, 1984). Even adults sometimes use counting-based procedures rather than retrieval to solve simple addition problems (Geary & Wiley, 1991; LeFevre, Sadesky, & Bisanz, 1994; Thibodeau, 1993), and very young children occasionally use retrieval rather than counting (Siegler & Robinson, 1982). Third, many procedures are developed without direct instruction as children solve problems in and out of school (Bisanz et al., in press; Ginsburg, 1986).

Recognizing that individuals have a repertoire of procedures that can be used adaptively has several important implications for developmental theory and for methods of assessment (Bisanz & LeFevre, 1990; Siegler & Shrager, 1984), two of which are of immediate relevance. One implication is that a level or stage of development is not necessarily defined most accurately in terms of the degree to which a single, dominant procedure is used. In mental arithmetic, it would be erroneous to assume that children at developmental level p use procedure x exclusively, that children at level p+1 use procedure y, and so on. The notion that stages or phases of development are defined in terms of single strategies, operations, or concepts has dominated developmental theory for a long time, but clearly this approach can be misleading.

Recognizing that children use diverse procedures to solve problems also implies that if we are to understand developmental change and sources of individual differences in performance, then we need to understand the process by which these procedures are selected. Considerable progress has been made on this issue over the last decade (Ashcraft, 1987; Siegler, 1987, 1988; Siegler & Shipley, in press; Siegler & Shrager, 1984), and a consensual view is emerging that appears to have broad applicability. Here we highlight some of the main characteristics of this emerging view by briefly describing elements of Siegler's (1988; Siegler & Shrager, 1984) theory of the selection process.

Selecting procedures. In Siegler's view, solution of arithmetic problems is a two-phase process (see Figure 2). When presented with a problem, a person first attempts to retrieve an answer directly from his or her memory for arithmetic facts. If that attempt fails, a second phase is engaged in which a backup procedure, such as *min* or *sum*, is used to generate or construct an answer. Thus whether a person uses retrieval or a backup procedure depends on the success of the retrieval phase, which in turn depends largely on knowledge of arithmetic facts.

In Siegler's model, knowledge of arithmetic facts is represented in terms of associative strengths between specific problems (e.g., 4+5) and possible answers (e.g., 6, 7, 8, 9, etc.). Retrieval of an answer from memory is probabilistic and depends on associative strength or activation levels [$Act(A_N)$], so that weakly associated answers to a problem are retrieved infrequently and strongly associated answers are retrieved often. Thus knowledge of arithmetic facts is represented in terms of *distributions of associative strengths* for different problem-answer combinations. Two such hypothetical distributions are illustrated

in Figure 3. Note that for 5+1, the association with one answer (6) is very strong and associations with other answers are weak, whereas for 4+5 none of the candidate answers is associated strongly. Such distributions are characterized by their degree of *peakedness*: the distribution of associative strengths for 5+1 is *peaked*, whereas the distribution for 4+5 is *flat*.

Distributions of associative strengths have an important influence on the selection process. Note in Figure 2 that an individual first encodes the problem and then sets a search parameter (n) that determines how persistent he or she will be in attempting to retrieve a solution. On the first attempt, if the retrieved answer reaches or exceeds a threshold called the *confidence criterion* (C), then the

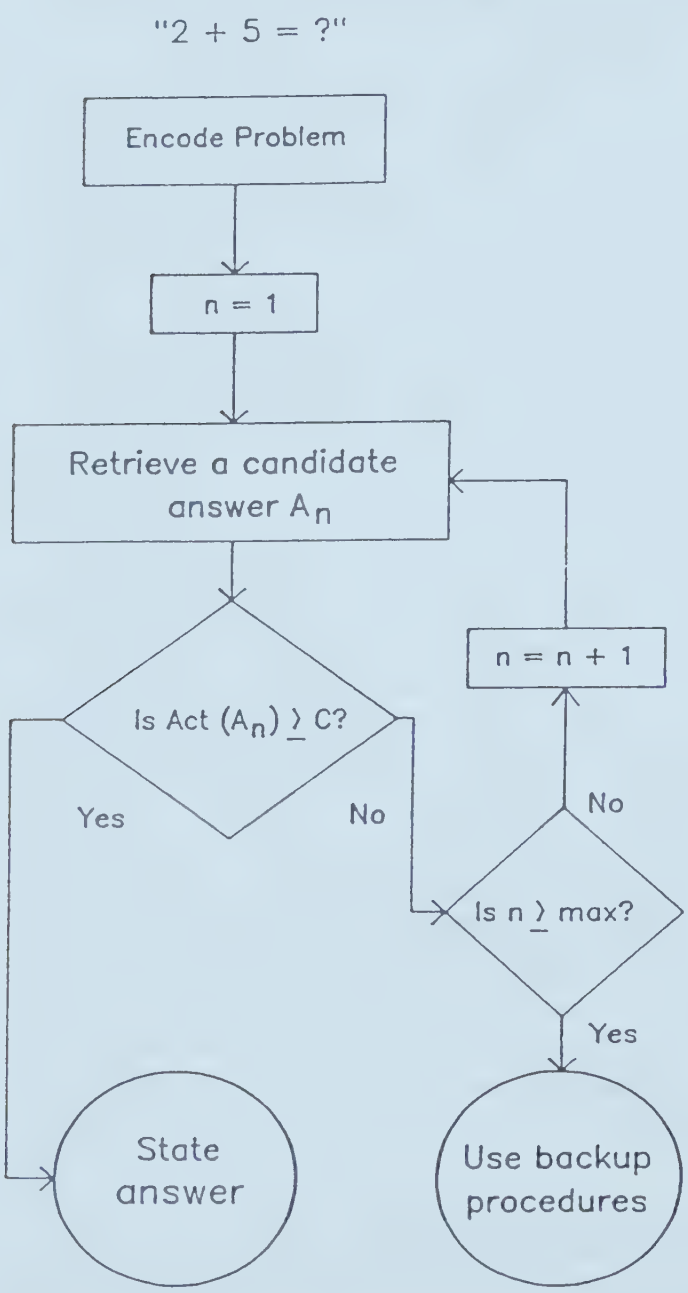


Figure 2. A simplified version of the selection process in Siegler's model of mental arithmetic. From "The information processing perspective on cognitive development in childhood and adolescence," by R. Kail & J. Bisanz, 1992, in R.J. Sternberg & C.A. Berg (Eds.), Intellectual Development *(p. 236). New York: Cambridge University Press. Copyright 1992 by Cambridge University Press. Reprinted with permission.*

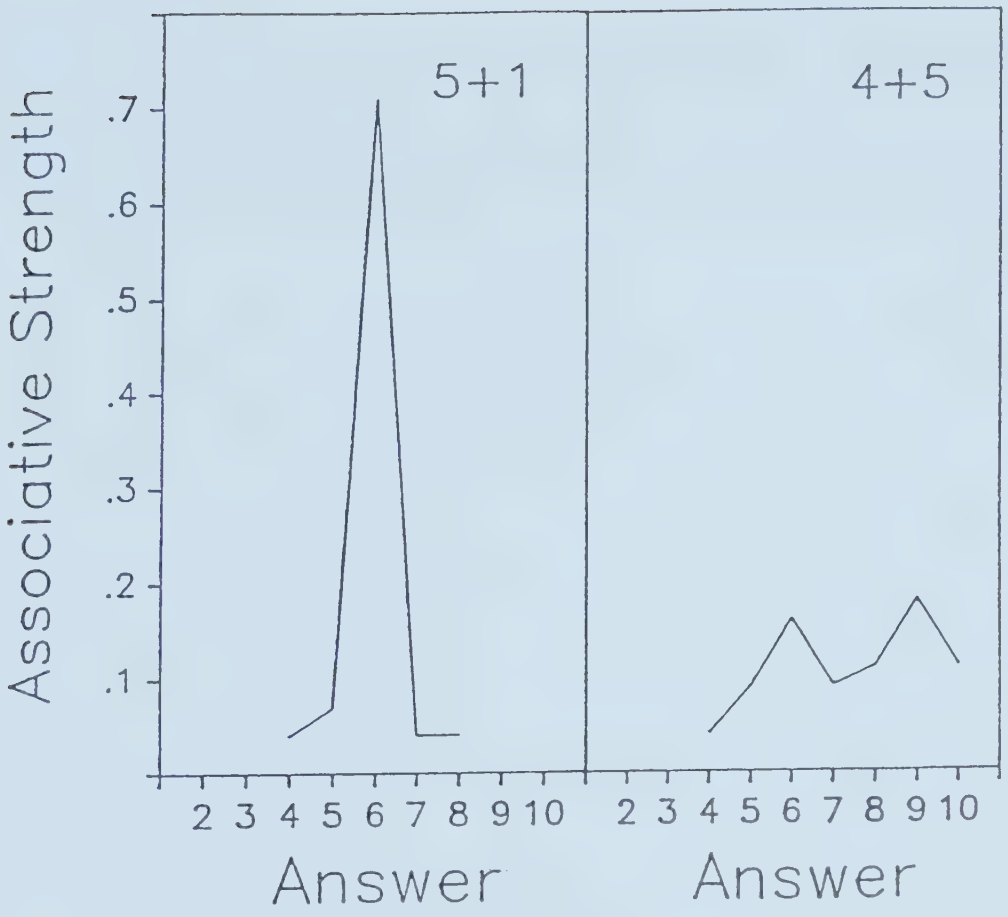


Figure 3. Distributions of associative strengths for two problems. From "Strategic and nonstrategic processing in the development of mathematical cognition" by J. Bisanz & J. LeFevre (1990) in D.F. Bjorklund (Ed.), Children's strategies: Contemporary views of cognitive development *(p. 220). Hillsdale, NJ: Lawrence Erlbaum Associates. Copyright 1990 by Lawrence Erlbaum Associates. Reprinted with permission.*

answer is stated. If the retrieved answer does not exceed the confidence criterion, then additional retrieval attempts are initiated until an answer exceeding the confidence criterion is retrieved or a limit (*max*) is reached on the number of retrievals or the time spent retrieving. If all retrieval attempts are unsuccessful, then a backup procedure is invoked to generate an answer. The success of the retrieval phase depends jointly on the distribution of associative strengths for a particular problem (i.e., how well the person knows the arithmetic fact in question) and on the level of the confidence criterion. For example, with the distributions illustrated in Figure 3 and a confidence criterion of .5, a child is likely to retrieve the answer to 5+1 and use a backup procedure to solve 4+5. With a confidence criterion of .1 a child is likely to retrieve answers to both problems, but accuracy on the second problem is likely to be low.

We hope this brief account conveys a sense of Siegler's theory of mental arithmetic and its development. Since its inception almost a decade ago (Siegler & Shrager, 1984), the theory has been successful in a number of senses (see Bisanz & LeFevre, 1990, and Siegler & Shipley, in press, for summaries). First, it helps us to account for a variety of empirical data, including correlations among solution latencies, accuracy rates, and the frequencies with which retrieval and backup procedures are used. Second, the model has been used to

generate novel and even counterintuitive predictions that have been supported empirically. Third, the model has been tested in mathematical and computer simulations that have led to interesting insights about the nature of learning mechanisms and, more generally, about cognitive development. Fourth, the model is generally compatible with more specific theories about retrieval mechanisms and problem solving in children and adults, and it has been useful for generating new research on these topics (e.g., Geary, Brown, & Samaranayake, 1991; LeFevre, Kulak, & Bisanz, 1991; LeFevre et al., 1994). Fifth, the model has considerable generality, and it has been used to identify certain similarities in development across domains (e.g., Siegler, 1986). Sixth, the theory has been revised successfully to account for additional phenomena, such as selection from among competing backup procedures (Siegler & Shipley, in press).

Finally, and of special importance for assessment, the model has proved useful for identifying insights about the nature of individual differences in performance (e.g., Geary et al., 1991; Siegler, 1988). For example, Siegler (1988) identified three groups of children who were classified as good students, not-so-good students, and perfectionists. Good students and perfectionists were more accurate than not-so-good students, but perfectionists used retrieval much less frequently than good students. These differences are readily interpretable in terms of two key parameters of Siegler's model: the peakedness of distributions and the level of the confidence criterion. Both perfectionists and good students have peaked distributions (as in the left panel of Figure 3), whereas the not-so-good students have flatter distributions. The difference between the two more successful groups is that perfectionists have higher confidence criteria than good students. The difference between good students and perfectionists is psychologically significant and potentially important for instruction, but it would have been impossible to detect with standard measures of achievement because both types of students would do well.

Understanding Arithmetic

Siegler's model concerns the computational processes used to solve arithmetic problems, but in its current form it sheds little light on the conceptual knowledge people have of arithmetic. Certainly how we solve problems is related to our understanding of arithmetic, but the nature of that relation is not very clear at present. The issue of understanding is central to many arithmetic teachers who commonly observe that their students are capable of solving simple problems when they are presented symbolically (e.g., 7-4=?) but fail miserably when required to interpret a situation arithmetically (e.g., "Erik has 7 tribbles, which is 4 more than Jason. How many does Jason have?"). Moreover, children who generate correct answers may differ considerably in their understanding of underlying arithmetic principles. Clearly the relation between what we understand about problems and how we solve those problems requires careful explication.

Developmental, cognitive, and instructional psychologists have begun to attack this issue, but progress has been difficult, in part because understanding and conceptual knowledge are often defined ambiguously or inconsistently. For example, conceptual knowledge has been described as "knowledge that is rich in relationships.... a connected web of knowledge, a network in which the

linking relationships are as prominent as the discrete pieces of information" (Hiebert & Lefevre, 1986, p. 3). As intuitively appealing as such definitions might be, they are difficult to operationalize for research and assessment purposes. Moreover, any single definition would tend to make understanding an all-or-none phenomenon and would obscure different forms of understanding that may emerge as new knowledge is acquired (Bisanz & LeFevre, 1992; Greeno, 1983).

One approach to this problem is to characterize the different forms of understanding that can be assessed and then to describe understanding in terms of performance across measures of these different forms. For example, Bisanz and LeFevre (1992) propose that contexts for assessing understanding can be organized in terms of two orthogonal aspects or dimensions, as illustrated in Table 1. The vertical aspect includes many different activities that can be used to assess forms of understanding. Consider the question of whether a child understands the principle of inversion in arithmetic (a+b-b must be equal to a). *Evaluation of procedures* refers to whether a student can recognize the validity of using a shortcut to solve problems such as 3+6-6 when that procedure is observed. These judgments help to identify whether students have some knowledge about the principle of inversion, even if they are incapable of executing an appropriate procedure or simply fail to do so spontaneously. These decisions are similar to judgments about grammaticality used by linguists and psycholinguists to determine whether individuals are sensitive to syntactic rules, whether or not those individuals are capable of stating the rule explicitly. *Application of procedures* refers to whether a student spontaneously uses a noncalculational shortcut to solve such problems as 3+6-6. Finally, *justification of procedures* is assessed by determining whether a student can describe the principle or concept that makes a procedure appropriate and legitimate, independently of whether the student uses the procedure spontaneously or recognizes the validity of the procedure correctly when observing it. Each of these "activities" is relevant to understanding, but none of them

Table 1
Contexts for Assessing Understanding

	Generality	
	Activity Narrow ——————> *Broad*	
Evaluation of procedures	Recognizing the validity of a procedure on one particular task	Making a similar judgment on a variety of related tasks
Application of procedures	Using an appropriate procedure spontaneously on one particular task	Using a similar procedure on a variety of related tasks
Justification of procedures	Describing the principle or concept that makes a procedure appropriate for one particular task	Generating a similar explanation on a variety of related tasks.

Note. From "Understanding elementary mathematics" by J. Bisanz & J. LeFevre, 1992, in J. Campbell (Ed.), *The nature and origins of mathematical skills* (p. 117). Amsterdam: Elsevier Science Publishers. Copyright 1992 by Elsevier Science Publishers. Reprinted with permission.

alone is sufficient for capturing a complete sense of what it means to understand an arithmetic principle.

The horizontal aspect of Table 1 is generality, which refers to the range of problems in which successful performance is observed. A common assumption in research on problem solving is that performance is more likely to reflect genuine understanding, as opposed to behavior acquired by rote, to the extent that it can be observed in a variety of new and/or different problems (e.g., Greeno, 1983). Carraher, Schliemann, and Carraher (1988), for example, found that construction foremen in Brazil were quite proficient at converting measurements from familiar scales on blueprints. For most of the foremen, however, performance dropped dramatically when they were required to use the same operations with unfamiliar scales, indicating that their understanding of ratio conversions was limited. In the case of understanding inversion, the question is whether students evaluate, apply, and/or justify an inversion-based shortcut on a variety of problems of different forms (e.g., b+a-b) or magnitudes (e.g., 3+6-6 versus 637+353-353).

We have two purposes in presenting Table 1. The first is simply to illustrate that cognitive, developmental, and instructional researchers have developed a variety of methods for assessing the extent to which a student "understands" arithmetic. The range of methods is impressive and informative. Second, any single measure of understanding is bound to provide an incomplete or misleading picture. An alternative is to conceive of understanding in terms of a profile in the contextual space defined by such aspects as activity and generality.

> Thus the issue for assessment of individuals is not one of deciding whether an individual has or does not have understanding, but rather one of determining the pattern of successes and failures in the contextual space. Similarly, the issue for studying acquisition is not how one group performs on a particular task as opposed to another group, but rather how understanding, or evidence for some form of understanding, "spreads" in this contextual space during the course of development or instruction. (Bisanz & LeFevre, 1992, p. 124)

Learning Environments

To this point we have discussed arithmetic as if the role of specific environmental or cultural influences is minimal. Traditionally, cognitive approaches tended to be focused on mental processes and their organization, paying little attention to the nature of the environments in which children learn. The cost of neglecting the role of the learning environment is obvious, however, and there is a growing body of research on factors related to culture (e.g., Stevenson & Stigler, 1992), schooling (Bisanz et al., in press) and home environments (Song & Ginsburg, 1987). For example, Stevenson, Lee, and Stigler (1986) initially documented striking differences in mathematics achievement among children in several countries on three continents. Many of these differences are apparent as early as grade 1 in nations where schooling at that level is universal, thus indicating that these differences are not due to selective sampling. In a second wave of research, Stevenson et al. (1990) and Stevenson, Chen, Lee, and Fuligni (1991) have begun to identify a large set of cultural and pedagogic differences among these nations that might contribute to these cross-national differences. Included are methods of instruction (e.g., the use of class discussion and

questioning), attitudes toward mathematics held by parents and teachers, and out-of-school practices (e.g., homework).

Because of constraints on space, we do not describe this research in any detail. It is important to recognize, however, that developmental and instructional research on the environments of learning mathematics is growing and making striking contributions to our understanding of acquisition. Psychologists are becoming increasingly aware that if we are to understand the development of academic skills, we need careful assessments of the environments in where those skills are acquired.

Research on the Development of Early Reading Skills

Developmental Frameworks

The need for a developmental framework for understanding how children learn to read was recognized by Chall (1979) in her seminal proposal about stages in reading acquisition. Chall's goal was to develop a Piaget-like theory of reading in which reading was seen as a form of problem solving. She outlined qualitative changes in reading as a problem solving activity over the course of schooling. Early stages were focused on relating print to speech and later stages on relating print to ideas. This shift leaves her framework open to criticisms that the proposed stages reflect changing emphases that exist in the elementary curriculum, rather than reflecting the ways in which children actually learn to read.

In contrast, information processing accounts of the development of reading have been enhanced by models of skilled reading in adults, as developed by

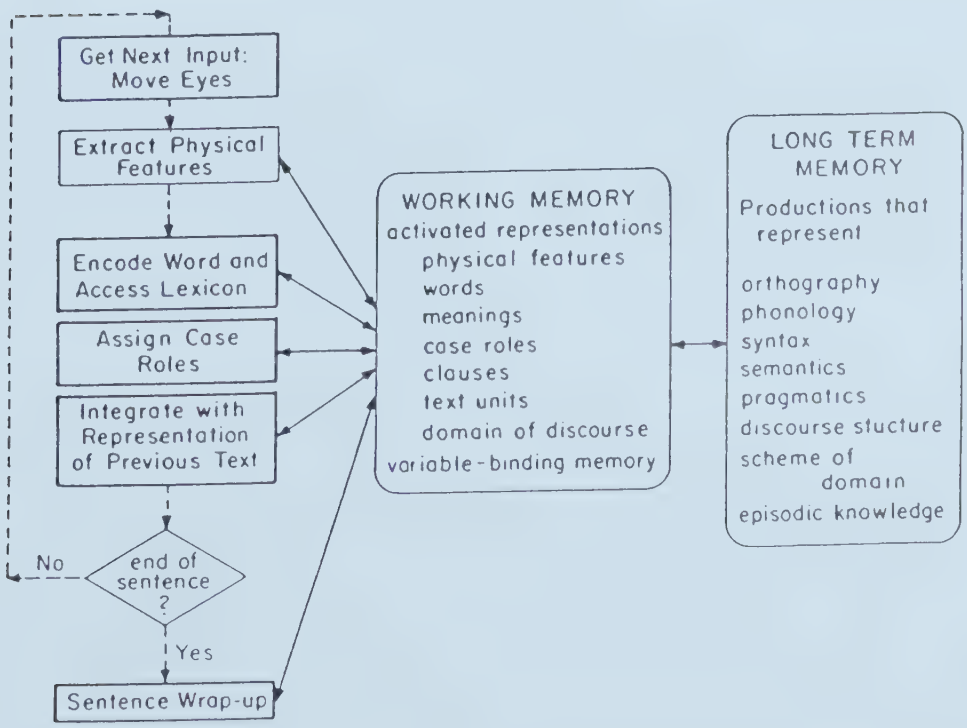


Figure 4. A schematic diagram of the major processes and structures in reading comprehension. From "A Theory of Reading: From Eye Fixations to Comprehension," by M.A. Just and P.A. Carpenter, 1980, Psychological Review, 87, p. 331. Copyright 1980 by the American Psychological Association. Reprinted with permission.

cognitive psychologists. Just and Carpenter's (1980) model of the cognitive processes that predict the eye fixations of skilled readers can serve as a vehicle to illustrate the major types of processes and structures that must be encompassed in any information processing account of skilled performance (see Figure 4). At least three questions must be addressed: What types of knowledge must the reader have in long-term memory? What is the role of working memory or limited attentional resources in performance? What are the major processes in reading and how do they interact given the reader's goal and the type of passage? From an information processing perspective, "reading" can be viewed as a family of related skills where processes could differ in temporal duration or organization or both, depending on the reader's goal and the nature of the reading materials.

For familiar words skilled readers have a large store of word codes referred to as a print lexicon by psychologists and a sight vocabulary by educators. Access to these codes involves fast and automatic retrieval, a process that uses few or no attentional resources. For novel words, skilled readers use a set of decoding or "back-up" strategies (Perfetti, 1992), including identification of subsyllabic units important in oral speech (e.g., Treiman, 1992), phonetic strategies that involve one-to-one mappings of individual phonemes to graphemes, and more sophisticated orthographic strategies that take advantage of the complexity of English spelling. These strategies all take advantage of the relations between oral language and print afforded by use of the alphabet in English. There is also some suggestion that some of the earliest and most ephemeral strategies that children use may involve the use of visual cues, treating English as if it were a logography like Chinese (e.g., Frith, 1985; Gough & Hillinger, 1980).

Children who are just learning to read in their native language begin with a relatively large oral lexicon and a nonexistent print lexicon. The critical question "What develops in early reading?" has multiple answers (see Perfetti, 1992). For familiar words it is both the number and nature of word codes, including the precision and redundancy of their representation in memory. Speed of accessing word codes also increases, but this change could well be a function of other developments. For unfamiliar words, what develops is the sophistication of back-up strategies.

Psychologists working in the area of reading and oral language hold the view that reading "piggybacks" on language in a complex way (Liberman & Shankweiler, 1979). Perfetti (1985) proposed a distinction between language processes that constitute a general reading ability and other processes of language and thought that occur in a variety of activities, including reading. Visual word recognition, for example, is a process unique to reading ability. This proposal enables one to ask whether some children could have good thinking and oral language skills but be poor readers. When asked this question, teachers generally agree some children fit this pattern.

Word recognition and thinking may be related in performance, however, because slow, inefficient retrieval of word codes is thought to take away limited resources from comprehension processes (Perfetti, 1985). Such inefficiencies can result in compensatory processing, the use of sentence-level information in the text to assist word recognition. This verbal inefficiency

should be evident in a deteriorating ability to understand the overall meaning of a text (Perfetti, 1985; Stanovich, 1980).

Many psychologists who study children's reading from an information processing perspective suggest that the word codes used in reading are also used in spelling. One difference between the two skills is that greater precision and redundancy of representation is required for spelling (Frith, 1985). Consequently, there is a body of related work on the retrieval of word codes and the development of back-up strategies in spelling (e.g., Ehri, 1987; Frith, 1985; McClure & Varnhagen, 1992; Treiman, 1992; Varnhagen & Treiman, 1993). A working hypothesis is that programs in spelling and writing can be used to facilitate reading and vice versa, through the development of more precise and redundant word codes and teaching strategies that will enhance performance on both tasks (Ehri & Wilce, 1987). Of importance to research and assessment is that the hypothesized relation between these skills opens new avenues for exploring competence.

This emerging information processing framework on the development of reading has many parallels with the work on early arithmetic and just as many implications for assessment. For developmental psychologists and assessment specialists, important tasks include determining the size and organization of a child's print lexicon, as well as examining the nature and efficiency of word codes. In addition, information about compensatory processing and the range and sophistication of back-up strategies could become important for measuring cognitive growth and achievement.

Individual Differences

Beyond this general framework for measuring the growth of all children, developmental research has begun to provide a picture of children with special strengths and weaknesses that make them more or less likely to profit from classroom instruction. For example, some children appear to have normal oral language skills just before they enter school, but they perform poorly on special tasks requiring that they segment and manipulate subsyllabic units of oral language. These children tend to score poorly on tests of reading achievement at the end of grade 1 or 2. Identification of these phonological deficits has been called a scientific success story (Stanovich, 1987).

New and provocative longitudinal research conducted by Scarborough (1990) suggests that their problem is related to a deficiency in language skills that is evident in different forms at younger ages. For instance, at age 2 these same children show deficits in syntactic complexity that are not evident later at school entry. Not surprisingly, many new studies are focused on appropriate intervention strategies. This research exemplifies the benefits of close links between basic research and assessment in areas where advances are rapid and understanding is vitally important to the welfare of children.

Similarly, research has shown that "precocious readers," those children who arrive at the school door already knowing how to read, seem to be distinguished from their peers by interest in "cracking the alphabetic code" rather than by IQ or home background (Siegler, 1991). Teachers need to take advantage of the passions of young experts to promote their intellectual growth. Assessment specialists and researchers need to identify behaviors reflecting the special interests some children may have that cause them to excel

in reading, writing, arithmetic, science, and other subject matter areas, as well as methods for motivating children who lack those special interests.

Learning Processes in Home Environments

Although developmental psychologists interested in reading often start with a focus on the mind of the child, analyses extend to understanding factors that affect learning in the various environments where children grow and learn. For example, insights into the lives of precocious readers spring from new interests among developmental researchers in documenting processes in home environments that promote literacy. Research on emergent literacy in preschoolers is focused on the naive theories of reading and writing that children develop before coming to school (e.g., Bissex, 1980) and what parents of young children do at home to promote literacy.

Developmental researchers with an interest in emergent literacy owe much to the pioneering work of educators such as Clay (1972), who observed preschoolers' exploratory behavior related to reading and writing and recognized its importance to school instruction. Other early pioneers include language researchers Ninio and Bruner (1978), who went into a home to document the development of oral language in a mother-child pair and became interested in joint book reading. Case studies and ethnographic research have been supplemented by experimental research and intervention studies (e.g., Cornell, Senechal, & Broda, 1988; Senechal & Cornell, 1993; and Whitehurst et al., 1988) focused on the types of questions parents ask their children during joint book reading and on how these questions influence memory and the acquisition of literacy skills.

Teachers' knowledge of this body of research has done a great deal to change the early elementary curriculum from a focus on "reading readiness" to an approach where children's earliest theories about reading are recognized and the need to jointly enlist the talents of the child's teachers at home and at school to facilitate learning. The challenge for educational assessment is to find ways to help teachers develop tools that will allow them to tailor their programs to meet the needs of children with particular types of naive theories or home reading experiences. Tools are also needed to help parents and teachers assess the quality of the dialogues they have with children during joint activities such as book reading. Diagnostic tools for determining the quality of a dialogue would allow parents and teachers to understand what is most effective and when they have achieved it.

Effects of Classrooms and Curricula on Cognition

Several studies have helped to make many developmental psychologists sensitive to the effects of classroom practices and curricula on psychological processes. For example, Evans and Carr (1985) conducted an exemplary study on the relation between philosophies of teaching reading (in this case, language experience versus decoding-oriented basal reader approaches to teaching reading), classroom learning activities, and the cognitive skills children acquire. They measured reading achievement and also looked for sensible patterns of positive correlations between reading and basic skills in a variety of areas, a criterion for evaluation they refer to as cognitive coherence. They found that children in language experience classrooms actually spent less time in reading

and language activities than children in traditional classrooms, a reality that was reflected in the achievement test scores. Furthermore the language experience program was not cognitively coherent in that negative correlations were found between linguistic ability and reading skill in these classrooms. This study clearly demonstrates the importance of ongoing classroom activities and the potential irrelevance of a teacher's stated philosophy to patterns of cognitive skills that children acquire.

Similarly, Perfetti, Beck, Bell, and Hughes (1987) reported data from a longitudinal study showing that the patterns of acquisition for different phonemic skills varied with the type of reading instruction children received. These studies have made it increasingly clear that, at a minimum, relevant characteristics of both subjects and classroom activities are important to consider in evaluating in reading research.

The Effect of Culture on the Study of Human Development

Finally, environments relevant to the development of reading include not only home and school, but also properties of the orthography of children's native language and the customs of their culture and subculture. For example, writing systems differ with respect to the basic linguistic units (e.g., phonemes, syllables, morphemes, words) they represent. Comparative developmental studies have begun to explore the differential effects of orthography on children's early acquisition of reading skill (e.g., Caravolas & Bruck, 1993).

Similarly, work on the literacy events in the homes of children from different communities (e.g., Heath, 1982) has gone a long way toward sensitizing developmental researchers and educators to aspects of our multicultural society that may enhance or detract from a child's opportunities to profit from classroom instruction. Moreover, comparative work on reading achievement in different cultures (e.g., Stevenson et al., 1985) has led not only to analyses of differences in home and school environments, but also to insights on differences in shared cultural values that relate to schooling.

In developmental psychology, the scope of investigation has broadened to one where each culture can be viewed as a treatment in a broad, "natural" experiment. Documenting these "treatment effects" is critical for understanding what is common about basic psychological processes and what varies with cultural context. This view of human development implies that cross-national studies of achievement must be designed and interpreted carefully so that the influence of culture on the development of academic skills can be well understood.

Implications for Assessment

Given these brief, selective summaries, we highlight several specific implications for the development of new methods of assessment.

Specific Implications

A wide range of methods now exists for developing assessments focused on identifying processes, rather than just products. It is widely acknowledged that achievement tests designed to assess products of learning, rather than processes, are inherently limited in terms of their value for teachers and for students. Decades of cognitive research have confirmed what teachers have known all along: students sometimes get the right answers for the wrong reasons, and they

sometimes get the wrong answers for the right reasons. We have tried to demonstrate that a wide range of methods exists for successfully identifying the cognitive underpinnings of performance on academic tasks. What is needed now is careful research and development aimed at adapting these methods for classroom use.

Assessments of processes and knowledge need to be closely tied to cognitive developmental theories. If the goal is to identify cognitive processes involved in children's academic performance, then an adequate theory about those processes is essential for developing assessment methods and for interpreting the results of assessment. Siegler (1988), for example, has shown how an explicit theory can provide a critical basis for diagnosing and characterizing individual differences in mental arithmetic. Theories can be be important for designing appropriate instruction to compensate for children's misunderstandings (e.g., Cummins, 1991) or for identifying developmental milestones that can serve as informative signals about a student's level of performance. The sensitivity and utility of an assessment will be proportional to the power and accuracy of the theory on which it is based; as Lewin notes, "There is nothing so practical as a good theory."

Assessments must be designed to diagnose diverse as well as modal abilities and skills. There is a tendency to describe stages or levels of competence in terms of a single element—a strategy, procedure, rule, code, or concept—that dominates cognition during a given period of time. Recent research suggests a different view: At any point in development, individuals may have a repertoire of diverse elements, and elements from this repertoire may be selected adaptively depending on external conditions or internal factors. According to some views, this diversity is not only common, it is an essential characteristic of adaptive systems (e.g., Siegler, 1989). From an instructional point of view, it may be useful to know not only what the dominant element is, but also to know (a) whether other elements are available to the child for use under certain conditions and (b) the characteristics of processing that determine selection of elements (e.g., the peakedness of distributions or the level of the confidence criterion in arithmetic). Assessment methods need to be developed to detect not only modal elements, but also to elicit the range of different elements available.

Hands-on research with children is critical for developing and selecting test items. Traditionally, test designers have attempted to designate certain items as requiring certain types of processes, based on their intuitions about solution processes. For example, according to Bloom's (Bloom, Englehar, Furst, Hill, & Krathwahl, 1956) popular taxonomy, items can be identified as requiring distinct processes such as "knowledge," "comprehension," or "application." Research with children suggests that such an enterprise is likely to be misleading: Children often solve problems with unanticipated methods, and even simple solutions typically involve an interaction among several different types of processes (Bisanz, Bisanz, & LeFevre, 1984; Gierl, 1993). To avoid misleading and adultocentric views of how problems are solved, the mind of the child must become not only a central focus for developmental psychology, but also for assessment. Items must be tested with children using a full range of meth-

ods (e.g., latencies, self-reports, and accuracy rates) to identify the range of solutions children use.

Assessments need to be designed to measure special interests, motivations, and attitudes. To fully understand the source of individual differences, assessments need to be designed to assess the extent to which children enjoy certain tasks or domains. This information could prove extremely useful to teachers who wish to adapt instruction to the particular goals and needs of individual children.

Assessments are needed of the social and cultural processes that characterize children's learning environments. Research has demonstrated important influences of people and factors in the environments where children learn, yet assessment practices often proceed as if learning environments were irrelevant or unmeasurable. Assessments of the social and cultural processes that characterize children's learning environments would not only help to provide more complete explanations of a child's performance, but they also should help teachers determine the kinds of remedial instruction that are likely to be most effective with particular children.

Epilogue

Our review of research on cognitive development and its implications for assessment would be incomplete without two additional reflections. First, we have described research on cognitive development and research on assessment as if the flow of information between these two areas were unidirectional. Such a state of affairs would be far from ideal. The flow of information should be reciprocal: the results of developmental research should inform the development of assessment methods and practices, and the development and use of assessment devices should inform and challenge research on development. Second, if this reciprocity is to be achieved, some researchers and curriculum specialists need to acquire expertise in both domains. We believe that productive interactions between assessment and cognitive development may well depend on the degree to which people in these specialities collaborate to produce individuals with the skills and insights that can arise from an interest and expertise in both domains.

Acknowledgments

The authors gratefully acknowledge support from grants provided by the Social Sciences and Humanities Research Council and by the Natural Sciences and Engineering Research Council. Correspondence should be addressed to either author at the Centre for Research in Child Development, Department of Psychology, University of Alberta, Edmonton, AB, Canada T6G 2E9.

References

Ashcraft, M.H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2*, 213-236.

Ashcraft, M.H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C.J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 302-338). New York: Springer-Verlag.

Ashcraft, M.H., & Fierman, B.A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology, 33*, 216-234.

Bisanz, J., Bisanz, G.L., & LeFevre, J. (1984). Interpretation of instructions: A source of individual differences in analogical reasoning. *Intelligence, 8*, 161-177.

Bisanz, J., & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D.F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 213-244). Hillsdale, NJ: Erlbaum.

Bisanz, J., & LeFevre, J. (1992). Understanding elementary mathematics. In J. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113-136). Amsterdam: Elsevier.

Bisanz, J., Morrison, F.J., & Dunn, M. (in press). Effects of age and schooling on the acquisition of elementary quantitative skill. *Developmental Psychology*.

Bissex, G.L. (1980). *GNYS at WRK*. Cambridge, MA: Harvard University Press.

Bloom, B.S., Englehar, M.D., Furst, E.J., Hill, W.H., & Krathwahl, D.R. (Eds.). (1956). *Taxonomy of educational objectives: Handbook 1. Cognitive domain*. New York: McKay.

Brown, J.S., & Burton, R.B. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*, 155-192.

Brown, J.S., & VanLehn, K. (1982). Towards a generative theory of "bugs." In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 117-135). Hillsdale, NJ: Erlbaum.

Cairns, R.B. (1983). The emergence of developmental psychology. In W. Kessen (Ed.) & P.H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 1. History, theory, and methods* (pp. 41-102). New York: Wiley.

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children's phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*, 1-30.

Carraher, T.N., Schliemann, A.D., & Carraher, D.W. (1988). Mathematical concepts in everyday life. In G.B. Saxe & M. Gearheart (Eds.), *Children's mathematics: New directions for child development* (vol. 41, pp. 71-88). San Francisco, CA: Jossey-Bass.

Case, R. (1993, March). Central conceptual structures and their manifestations in specific task performance. In D. Kuhn (Chair), *Identifying the general in the increasingly context- and domain-specific study of cognitive development*. Symposium conducted at the meeting of the Society for Research in Child Development, New Orleans.

Chall, J. (1979). The great debate: Ten year later, with a modest proposal for reading stages. In L.B. Resnick & P.A. Weaver (Eds.), *Theory and practice of early reading Vol. 1* (pp. 29-55). Hillsdale, NJ: Erlbaum.

Clay, M.M. (1972). *Reading: The patterning of complex behavior*. Auckland, NZ: Heinemann Educational.

Cooney, J.B., & Ladd, S.F. (1992). The influence of verbal protocol methods on children's mental computation. *Learning and individual differences, 4*, 237-257.

Cornell, E.H., Senechal, M., & Broda, L.S. (1988). Recall of picture books by 3-year-old children: Repetition and testing effects in joint reading activities. *Journal of Educational Psychology, 80*, 537-542.

Cummins, D.D. (1991). Children's interpretation of arithmetic word problems. *Cognition and Instruction, 8*, 261-289.

Ehri, L.C. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*, 5-31.

Ehri, L.C., & Wilce, L.S. (1987). Does learning to spell help beginning readers learn to read words? *Reading Research Quarterly, 18*, 47-65.

Evans, M.E., & Carr, T.H. (1985). Cognitive abilities, conditions of learning, and the early development of reading skill. *Reading Research Quarterly, 20*, 327-350.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J.C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (pp. 301-330). Hillsdale, NJ: Erlbaum.

Geary, D.C., Brown, S., & Samaranayake, V.A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787-797.

Geary, D.C., & Wiley, J.G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 6*, 474-483.

Gierl, M.J. (1993). *Evaluating Bloom's cognitive levels in the table of specifications for a grade 6 mathematics achievement test*. Unpublished master's thesis, University of Alberta.

Ginsburg, H.P. (1986). Academic diagnosis: Contributions from developmental psychology. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 235-260). New York: Plenum.

Gough, P.B. & Hillinger, M.L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society, 30*, 180-196.

Greeno, J.G. (1983). Forms of understanding in mathematical problem solving. In S.G. Paris, G.M. Olson, & H.W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 83-111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Heath, S.B. (1982). What no bedtime story means: Narrative skills at home and school. *Language and Society, 11*, 49-75.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Erlbaum.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490-501.

Kail, R., & Bisanz, J. (1982). Cognitive strategies. In C.R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 229-255). New York: Academic Press.

Kail, R., & Bisanz, J. (1992). The information-processing perspective on cognitive development in childhood and adolescence. In R.J. Sternberg & C.A. Berg (Eds.), *Intellectual development* (pp. 229-260). New York: Cambridge University Press.

Kail, R.V., Jr., & Hagen, J.W. (Eds.) (1977). *Perspectives on the development of memory and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

LeFevre, J., Kulak, A.G., & Bisanz, J. (1991). Individual differences and developmental change in the associative relations among numbers. *Journal of Experimental Child Psychology, 52*, 256-274.

LeFevre, J., Sadesky, G., & Bisanz, J. (1994). *The role of strategy selection in mental addition: Reassessing the problem-size effect in adults*. Unpublished manuscript.

Liberman, I.Y., & Shankweiler, D. (1979). Speech, the alphabet, and teaching to read. In L.B. Resnick & P.A. Weaver (Eds.), *Theory and practice of early reading* (vol. 2, pp. 109-132). Hillsdale, NJ: Erlbaum.

McClure, K., & Varnhagen, C.K. (1992). *Do children spell the way they think they do? Predicted versus reported strategy use by early elementary school children*. Paper presented at the meetings of the American Educational Research Association, San Francisco.

Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labeling. *Journal of Child Language, 5*, 1-14.

Pachella, R.G. (1974). The interpretation of reaction time in information- processing research. In B.H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.

Perfetti, C.A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C.A. (1992). The representation problem in reading acquisition. In P.B. Gough, Linnea C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Erlbaum.

Perfetti, C.A., Beck, I., Bell, L.C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.

Royer, J.M., Cisero, C.A., & Carlo, M.S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research, 63*, 201-243.

Russo, J.E., Johnson, E.J., & Stephens, D.L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*, 759-769.

Scarborough, H.S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728-1743.

Senechal, M., & Cornell, E.H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*, 361-374.

Siegler, R.S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481-520.

Siegler, R.S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46* (Whole No. 189).

Siegler, R.S. (1986). Unities across domains in children's strategy choices. In M. Perlmutter (Ed.), *Minnesota symposium on child development* (pp. 1-48). Hillsdale, NJ: Erlbaum.

Siegler, R.S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250-264.

Siegler, R.S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59*, 833-851.

Siegler, R.S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology, 40*, 353-379.

Siegler, R. (1991). *Children's thinking* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Siegler, R.S., & Robinson, M. (1982). The development of numerical understanding. In H.W. Reese & L.P. Lipsitt (Eds.), *Advances in child development and behavior* (vol. 16, pp. 241-312). New York: Academic Press.

Siegler, R.S., & Shipley, C. (in press). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Erlbaum.

Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.

Simon, H.A. (1962). An information processing theory of intellectual development. *Monographs of the Society for Research in Child Development, 27* (Serial No. 82).

Song, M.J., & Ginsburg, H.P. (1987). The development of informal and formal mathematical thinking in Korean and U.S. children. *Child Development, 58*, 1286-1296.

Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of fluency in reading. *Reading Research Quarterly, 16*, 32-71.

Stanovich, K.E. (1987). Introduction. *Merrill-Palmer Quarterly, 33*(3).

Stevenson, H.W. (1972). *Children's learning*. New York: Appleton-Century-Crofts.

Stevenson, H.W., Chen, C., Lee, S.Y., & Fuligni, A.J. (1991). Schooling, culture, and cognitive development. In L. Okagaki & R.J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 243-268). Hillsdale, NJ: Erlbaum.

Stevenson, H.W., Lee, S.Y., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the United States. *Child Development, 61*, 1053-1066.

Stevenson, H.W., Lee, S.Y., & Stigler, J.W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231*, 693-699.

Stevenson, H.W., & Stigler, J.W. (1992). *The learning gap*. New York: Summit Books.

Stevenson, H.W., Stigler, J.W., Lee, S.Y., Lucker, G.W., Kitamura, S., & Chen-chin, H. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56*, 718-734.

Thibodeau, M.H. (1993). *Automaticity of multiplication facts and procedure use for multiplication problems: Evidence for ADAMM*. Unpublished master's thesis, University of Alberta.

Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 65-106). Hillsdale, NJ: Erlbaum.

Varnhagen, C.K., & Treiman, R. (1993). *On-line investigation of children's spelling*. Paper presented at the meetings of the American Educational Research Association, Atlanta.

Werner, H. (1937). Process and achievement. *Harvard Educational Review, 7*, 353-368.

Wertsch, J.V., & Kanner, B.G. (1992). A sociocultural approach to intellectual development. In R.J. Sternberg & C. Berg (Eds.), *Intellectual development* (pp. 328-349). New York: Cambridge University Press.

Whitehurst, G.J., Falco, F.L., Lonigan, C.J., Fischel, J.E., DeBaryshe, B.D., Valdez-Menchaca, M.C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24*, 522-559.

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 147-161*

John R. Kirby

and

Rosamund A. Woodhouse
Queen's University

Measuring and Predicting Depth of Processing in Learning

Recent research regarding depth of processing in student learning is reviewed. The nature of processing depth is described as a dispositional construct and as a characteristic of task-specific strategies. Efforts to measure processing depth, as shown in students' text summaries, are reviewed. Measures include content scores (based on a weighted sum of propositions included), depth ratings, and analyses of subjects' verbalizations during study. The individual difference and task variables associated with processing depth are considered, and a number of experimental studies reviewed. The article concludes by considering several ways in which depth may be measured in educational assessment.

Dans cet article, on examine les recherches récentes vis-à-vis le degré de l'approfondissement de la matière dans le contexte de l'apprentissage chez l'élève. La nature de l'approfondissement de la matière est considérée comme étant une construction intellectuelle et dispositionnelle et comme une caractéristique des stratégies centrées sur des tâches et des habiletés spécifiques chez l'élève. On examine aussi dans le contexte d'analyses sommaires des élèves comment mesurer le degré de l'approfondissement de la matière. Les éléments qui aideront à déterminer ce degré seront le contenu des scores (basé sur la somme totale de ce qui est proposé), l'évaluation du degré de l'approfondissement de la matière, et les analyses des expressions verbales qu'utilisent les élèves pendant l'étude. On considère aussi la différence individuelle ainsi que les variables de chaque tâche associées au degré de l'approfondissement de la matière. Certaines études expérimentales qui traitent de ce sujet sont examinées. On conclu l'article en considérant différentes façons permettant de déterminer le degré de l'approfondissement de la matière quand il s'agit d'évaluation en éducation.

It has long been argued, in both education and psychology, that *depth of processing* is valuable in learning. Although definitions have varied or been lacking, there is general agreement that characteristics of greater depth include: focus on main ideas rather than details, greater connection of new information with prior knowledge, less reliance on verbatim recall, reordering or drawing connections between disparate sections of a text, drawing inferences or general principles, and ability to apply learning to a novel task. These characteristics have *going beyond the information given* as their basic feature. Depth of process-

John Kirby is a professor of education. After his doctoral work in educational psychology at the University of Alberta, he worked for 11 years in the Faculty of Education at the University of Newcastle, NSW, Australia. His research interests include learning and studying strategies, the cognitive processes of comprehension, and learning disabilities.
Rosamund Woodhouse is a research associate at Queen's University, where she completed her doctorate in psychology in 1990. Her research interests include learning and studying strategies and the cognitive processes of bilingualism.

ing is thought to contribute to the quality of learning by enhancing recall of information, increasing the level of importance of the information recalled, and allowing the learner to generalize learning.

The basic notion of depth of processing is that much "new" information is relatively raw, uninterpreted, and unintegrated, and needs to be transformed, interpreted, and related to other information before it can be used optimally. This begs the question of how the information is to be used, that is, what the task is. In Western education, much lip service and some serious attention is paid to the proposition that learning and teaching should focus on generalizable principles, problem solving, and understanding. Teachers and students alike show disrespect for learning and teaching that emphasize "just the facts," are not applied to "real" problems, are "low level," or require "regurgitation." In spite of these espoused beliefs, much teaching and learning is shallow, and there is legitimate concern that this is the result of evaluation practices and perceptions of them.

If beliefs about the value of depth are to be tested, and, if found to be valid, applied, we need to examine what depth is, how it can be measured, and what factors encourage it. The purposes of this article are to (a) examine the concept of depth of processing, (b) develop criteria for measuring depth of processing, (c) consider the personal and situational factors that predict depth of processing, and (d) review recent studies of learning from text that bear on these issues. This leads to discussion of methods of addressing depth of processing in educational assessment, and to the implications of a shift to this goal of assessment.

Depth of Processing

Many of our notions of depth originate in memory research in experimental psychology. In 1972, Craik and Lockhart proposed a levels of processing frame-

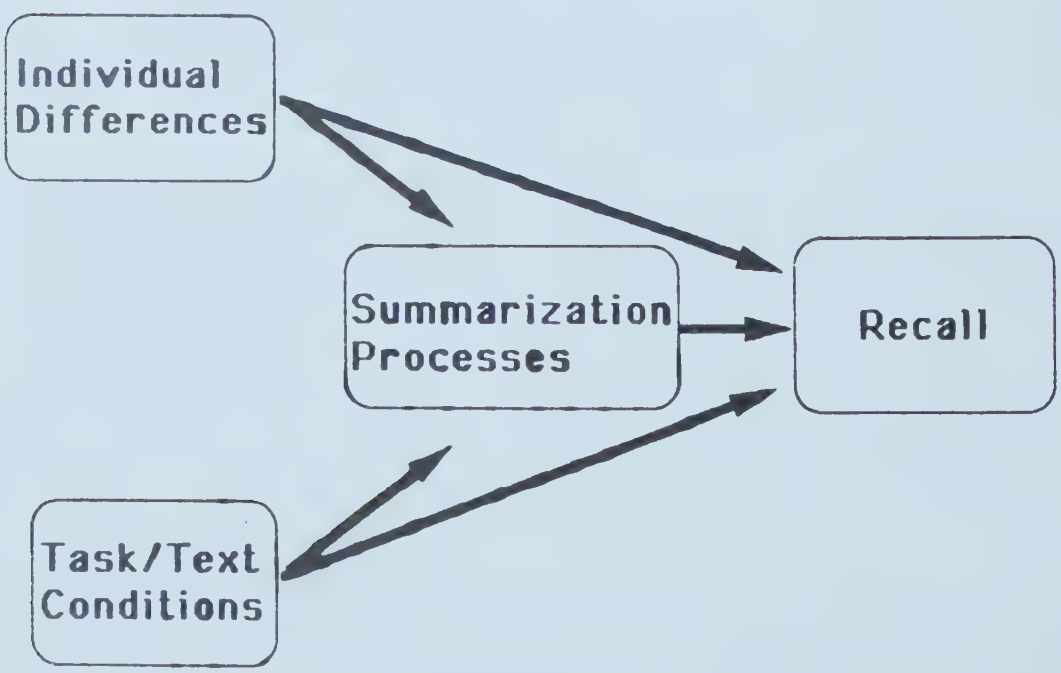


Figure 1. Model showing effects of individual differences, task factors, and study processes on recall.

work for understanding memory, where the different levels represented degrees of depth. The three levels so distinguished, the physical, phonemic, and semantic, were greatly different from each other. Only the semantic level is critical to the focus of this article, meaningful student learning. We suggest that there are degrees of depth in semantic processing that are associated with both the quantity and quality of what is remembered.

Consider Figure 1, adapted from similar figures by Biggs (1994), Duncan and Biddle (1974) and Biggs and Kirby (1984). This path diagram shows that individual differences and situational (task) factors affect processing during learning, and that all three of these affect subsequent learning outcomes. Depth of processing may be associated with each of these four components. For example, students may vary in the degree to which they are inclined toward deeper processing (an individual difference variable), which may in turn be due to other individual differences (in prior knowledge, verbal ability, etc.). As the memory researchers have shown, situations exert considerable influence over processing; task manipulations such as insertion of deeper questions, or provision of an advance organizer, are intended to increase depth of processing. Depth of processing may also be seen as something controlled by the individual student's motives and strategies at a particular time, whether during study (the middle box of Figure 1) or at performance (the rightmost box). Each of these components of depth poses problems for measurement. We review each of the components in turn, describing some recent research results, and indicating what the measurement issues are.

Disposition Toward Depth of Processing

Depth of processing has been considered as a disposition, as an individual difference regarding the likelihood of engaging in deeper processing (e.g., Biggs, 1987; Entwistle, 1988). In this sense it has been measured either by questionnaire or interview, and has been seen as a stable, transsituational characteristic, not unlike a personality trait or cognitive style. Biggs (1987) defined three dimensions of students' "approaches to learning": Deep, Surface, and Achieving, each having a motive and a strategy component. The Deep Approach to learning is characterized by intrinsic motivation and meaningful learning strategies, the Surface Approach by extrinsic motivation and rote learning strategies, and the Achieving Approach by achievement motivation and task-oriented, organized strategies. Students may have any combination of these characteristics; thus, for example, a student may show high levels of both Deep and Surface Approaches. This is not necessarily inconsistent, as some learning may require both; in fact, a combination of Deep and Surface is typical of university Science students.

There is considerable evidence that these approaches to learning variables are related to student learning, but the effects are far from powerful. In general, the surface approach is negatively related to success, the deep and achieving approaches positively related. Overall, the approach variables usually account for less than 15% of the performance variance. Biggs and Rihn (1984) found that university students asking for help at a learning assistance center had a predominantly surface approach, which shifted toward a deeper approach after completion of the center's program.

Although no one would seriously suggest that these approach measures be used to assess learning outcomes, they may be worth considering in conjunction with other factors. For example, Lai and Biggs (1994) have shown that mastery learning conditions favor, and are preferred by, high school students who report a surface approach to learning; although mastery learning improved group achievement, this was due to improvements by the surface learners at the expense of a deterioration in the performance of the deeper learners. As we see in the next section, students' approaches to learning also may determine how they react to instruction designed to increase depth of processing. These measures may have some benefit in helping students understand their own nature as learners. The validity of the transsituational nature of these approach variables should also be further examined (Biggs, 1993).

It is important to determine the antecedents of these approach variables. For example, a deep approach may require greater verbal ability and available working memory. A deep approach may only develop in an encouraging context; alternatively, the surface approach may develop as a response to particular experiences. The existing research tends to deal with approaches as already existing, rather than inquiring how they developed (see Biggs, 1987, 1993). If different educational tasks require approaches that differ in the relative amounts of deep and surface approaches, then optimal performance should be shown by students who understand these contingencies and can adapt their approach to the task at hand.

Conditions Affecting Depth of Processing

Depth of processing is a construct that is far from easy to measure unambiguously. A good experimental strategy for dealing with this is to measure the results of the process rather than attempting to measure the process itself. If the hypothesis is that greater depth will result in enhanced learning, design two treatments that differ in the depth they encourage, and observe the outcomes. There have been numerous attempts to do this with respect to constructs related to depth of processing. For the purposes of the present article, we describe two techniques, text absent summarization and poorly structured text.

Text absent summarization. Many instructors fear that students are too passive when they study. This passivity allows a form of comprehension that does not require much depth of processing and may result in shallow or fragile learning. Hidi and Anderson (1986) proposed that students would be forced to become more active if they were asked to summarize the text they were studying, but only after the text had been removed. Text absent summarization should encourage students, or at least some of them, to be more integrative and constructive during study, which should result in greater subsequent recall of important information from the text. We have examined this technique in a number of studies (Kirby & Pedwell, 1991; Stein & Kirby, 1992; Woodhouse & Kirby, 1991).

In general, we have not found across-the-board improvements in learning as a result of text absent summarization, in comparison with more traditional text present summarization. Instead, text absent summarization appears to benefit certain students, while adversely affecting the performance of others. The students who benefit are those of greater initial ability, those who write better summaries, or those who report a deep approach to learning. Text absent

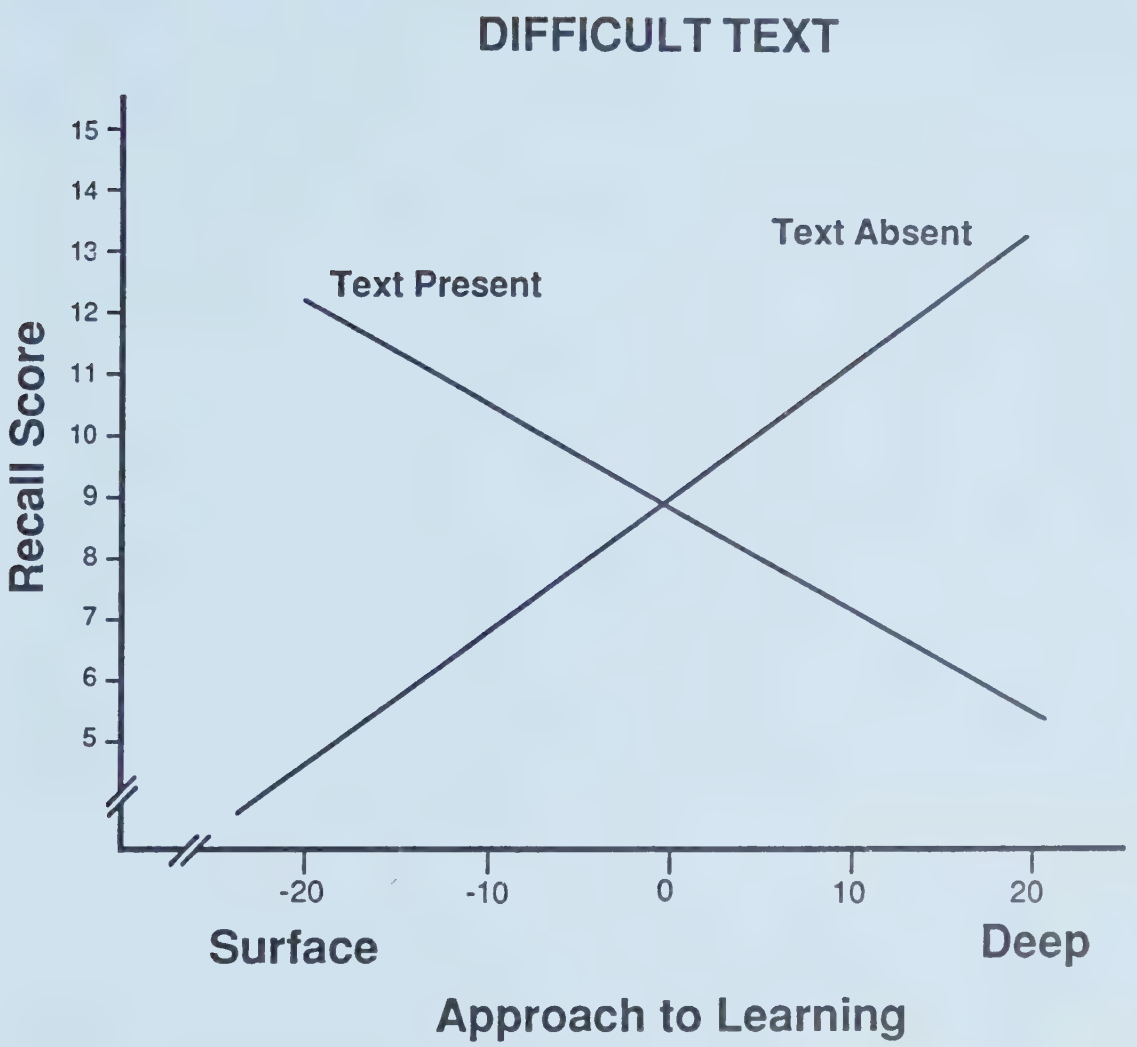


Figure 2. Regression lines illustrating interaction between task condition and students' approaches to learning (adapted from Kirby & Pedwell, 1991).

summarization is essentially a more onerous or stressful form of study: if you can cope with its demands, it helps; otherwise it may be counterproductive.

Kirby and Pedwell (1991) found an interaction between summarization condition (text absent/present) and university students' reported approaches to learning (defined as the relative tendency toward a deep or surface approach, calculated by subtracting the surface score from the deep score, see Figure 2). In the text absent condition, recall was positively associated with the relative depth of the students' habitual approach to learning; in the text present condition, this relationship was negative.

Stein and Kirby (1992) employed the text absent/present technique with grade 7 students. They found that text absent subjects wrote deeper summaries and that those who did remembered more of the important information in the text. We return to the analysis of summary quality after considering a second technique for enhancing depth of processing.

Poorly structured text. A second comparison has been made between text that is well structured (user-friendly or considerate) as opposed to text that is poorly structured (less friendly or considerate). Although very poorly structured text (sentences or words in random order) would be impossible to com-

prehend, much less process deeply, poorly structured text that is merely difficult (i.e., still comprehensible but at greater effort) has been shown to enhance processing, at least in some subjects (Kintsch, 1990; Simpson, 1992). Our interpretation of this result is that poorly structured text leads some subjects to more active and deeper processing.

In a recent study (Simpson, 1992; Woodhouse, Kirby, Simpson, & Hadwin, 1992), we had university student subjects study one of two versions of a text, one in which the sequence of main ideas was maintained and one in which it was disrupted. Subjects were asked to summarize the text and two days later to do a free and a probed recall of it. Although the poorly structured group wrote less extensive summaries than the well-structured group, they did not differ in subsequent free recall. By itself, this is only weak support for the effects of poorly structured text. However, two other results offer stronger support. First, an interaction appeared between experimental condition and subjects' efficiency of access to working memory (as measured by a speed of word naming test); this interaction showed that poorly structured text yielded superior free recalls for subjects with more efficient working memory (see Figure 3). Presumably, more efficient working memory processes offered these subjects the opportunity to perform deeper processing while reading and summarizing the text. This was supported by the second finding, which was that the quality of summary

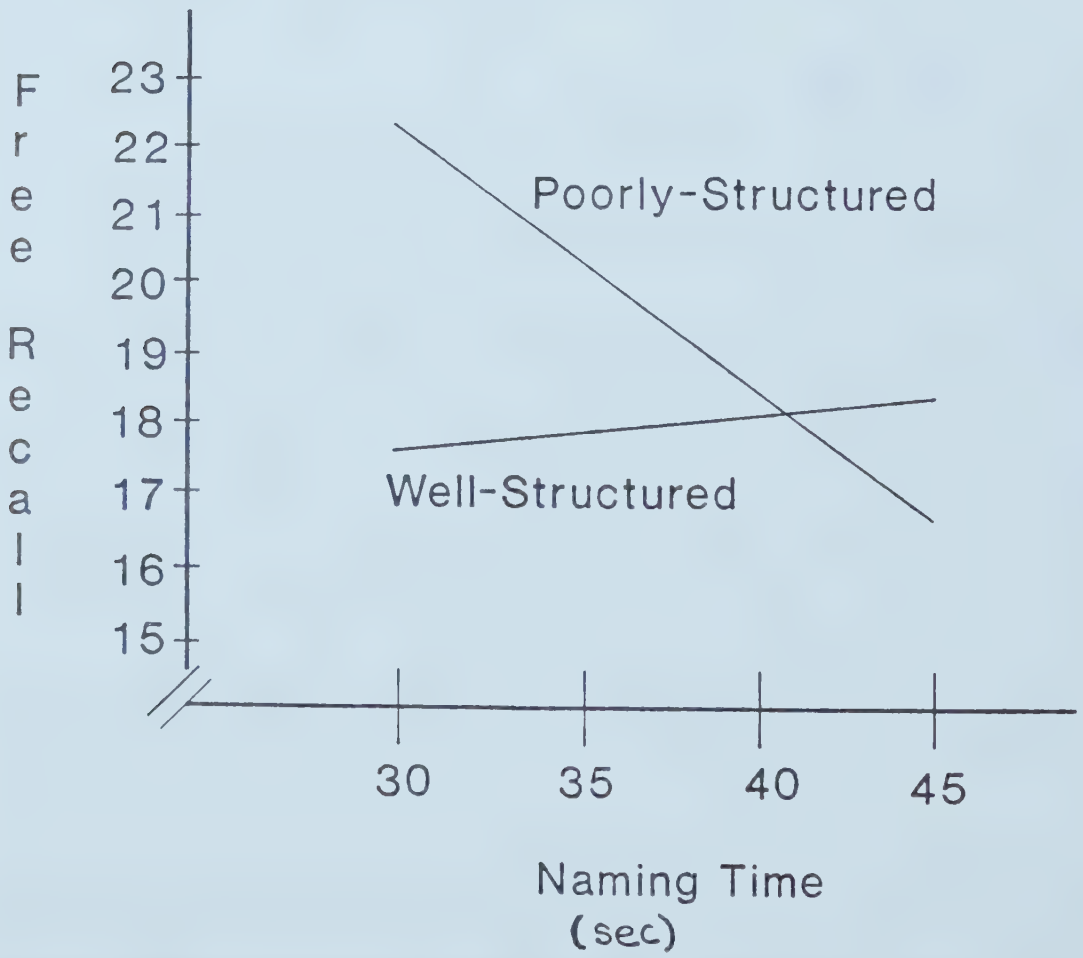


Figure 3. Regression lines showing interaction between working memory efficiency and text structure (adapted from Simpson, 1992).

written was a powerful predictor of recall for the poorly structured group. This latter effect eliminated the working memory effect, suggesting that greater working memory efficiency allowed deeper processing during study, which resulted first in a better summary, and second in better recall. (We examine the summary quality measures in the following section.) These effects appeared for free recall, but not for probed recall. In free recall, subjects must rely on their knowledge structures to "construct" a response, whereas in probed recall the question acts as a cue to trigger a response; in the latter case depth of processing is not as critical.

In a subsequent study (Peters, 1993), we replicated the working memory and text structure effects, and showed that instruction in deeper processing during study/summarization improved subjects' performance. We interpret this to mean that task conditions can have a powerful effect on the quality of processing employed by students during study, and that there is considerable scope for improvement. We return to these issues later in the article.

More generally, these task-imposed differences in processing offer several potential areas of application. First, they force us to recognize that tasks do differ in the processing they elicit and that these differences may not be constant across subjects. To accurately assess depth of processing, a range of task formats will need to be used, and the notion of a fixed "ability" for depth of processing may either be mistaken or in need of stated contextual conditions. Second, they suggest the possibility of a form of dynamic testing, in which some baseline level of processing is compared with that obtained in a task format that encourages deeper processing. Third, these different tasks may be valuable as training tasks to improve depth of processing. There is little danger of teaching to the test here, as the greater depth in study is exactly what we are attempting to encourage.

Depth of Processing During Study

To this point we have considered the first two boxes in Figure 1; essentially this is the same as aptitude-treatment interaction research, relating outcomes to some interaction between stable individual difference variables and environmental factors (more generally, Person-Task interactions). The model becomes more complex, and may gain more explanatory potential, with the addition of the Study Process component (Person-Task-Process interactions). Few doubt the importance of cognitive processes during learning; the question, rather, is how to measure these processes. We have used a variety of methods for measuring these processes during summarization.

Summarization seems an ideal context in which to examine study processes, because the student leaves a visible record of what he or she thinks is important. There are of course limitations. Summarization may be unfamiliar to some students, and some may not choose to write down what they think is important. Summarization may change the nature of learning, so that what we observe is not what would otherwise have occurred. Figure 4 presents a schematic outline of the cognitive processes involved in summarization. These processes are in three groups: on the left, those related to the student's comprehension of the text; in the middle, the summarization processes that depend on comprehension; and on the right, a description of the study product (the summary) that emerges. Comprehension processes are clearly relevant to

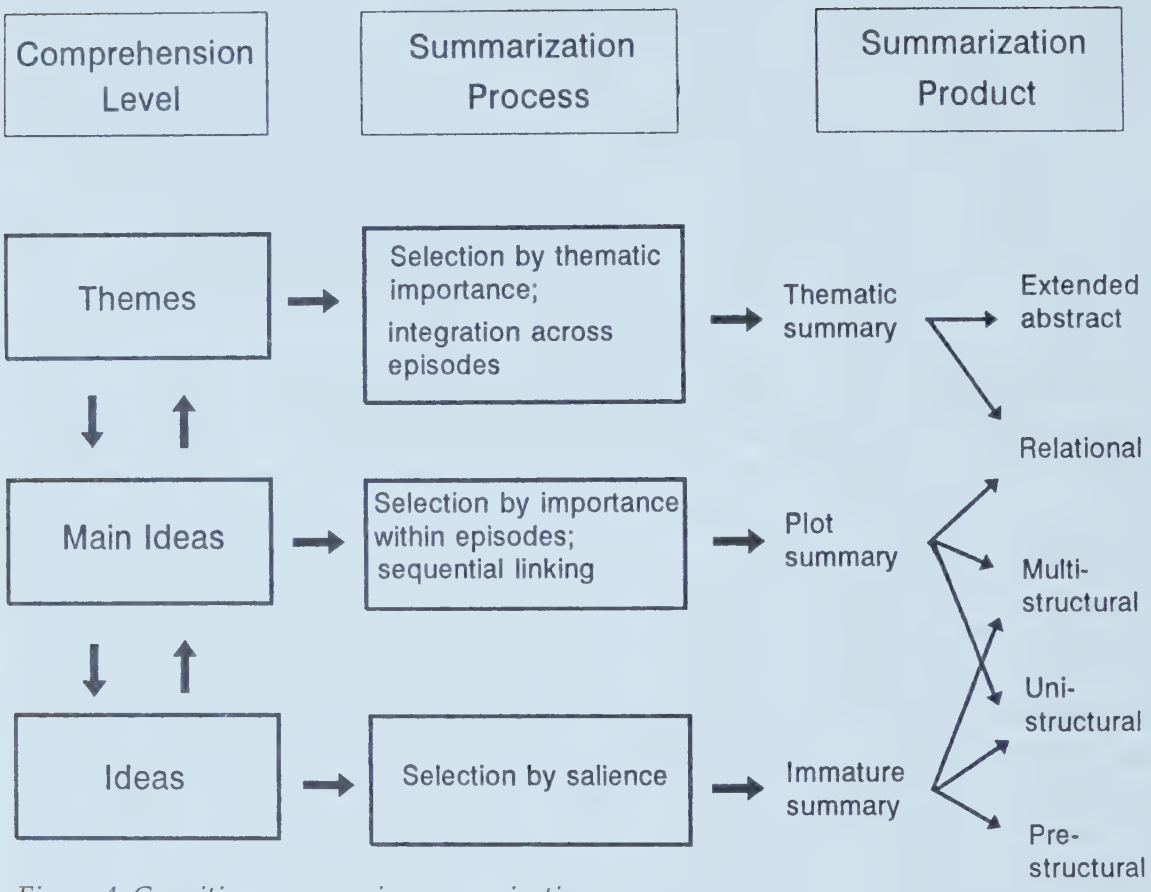


Figure 4. Cognitive processes in summarization.

learning from text and may be characterized in terms of the level at which they operate (Kirby, 1988). Three levels are shown, those relating to ideas (propositions), main ideas (macropropositions), and themes. Comprehension at the idea level is able to deal with individual ideas, as perhaps represented in a sentence, but not able to integrate across ideas or sentences or to recognize the relative importance of ideas. Comprehension at the main idea level can integrate across ideas or sentences and recognize relative importance of ideas, but does not go beyond the text to recognize the relative importance of main ideas. Comprehension at the thematic level does go beyond the text, applying an abstract frame of reference to its interpretation (see Kirby, 1988 and 1991 for more on this).

Summarization processes consist essentially of selection, linking, and construction. Selection is the means by which ideas are identified for inclusion in the summary. Linking is the process by which these ideas are connected, and construction is brought to bear when no appropriate idea can be selected or when an interpretation is made. These processes depend critically on the level of comprehension operating. If comprehension is operating at the idea level, then selection can only be on the basis of salience ("this is interesting") or by a rote rule ("copy the first sentence of each paragraph"); linking and construction hardly operate at all. If comprehension is at the level of main ideas, then selection can operate quite effectively in main idea units (roughly, paragraphs), but poorly across them. Linking works to form a sequential structure, essentially a plot line. Construction can operate in main idea units to form ideas, but

does not operate effectively between main idea units. If comprehension is at the thematic level, then selection, linking, and construction can operate effectively, with the lower levels interpreted in terms of the thematic level. Whereas main idea comprehension yields summaries that attempt to pay equal attention to each section or episode (until fatigue sets in or the paper runs out), thematic comprehension allows summaries that pay disproportionate attention to certain main ideas or even construct new main ideas.

These summary processes produce summaries of different quality. Idea level comprehension and selection by salience are unlikely to produce anything more than an immature summary. Main idea comprehension is more likely to result in a plot summary, an episode-by-episode recounting of the text. Thematic comprehension should result in a higher level, more abstract summary, which may in fact omit many of the details included at lower levels. These types of summaries can be characterized in terms of Biggs and Collis' (1982, 1989) SOLO (Structure of Observed Learning Outcomes) Taxonomy: immature summaries will either be prestructural (they miss the point), unistructural (they pick out one point), or at best multistructural (they pick out several points without integrating them). Plot summaries are essentially multistructural, but in some texts they may succeed in being relational (presenting a coherent framework within the bounds of the text). Thematic summaries will be at least relational, and should be able to reach Biggs and Collis' highest level, extended abstract (adoption of a frame of reference that is independent of the text). We return to the SOLO Taxonomy later, as this is one method of assessing summary depth.

We have used three different approaches to measuring these processes during summarization: measures of summary content and summary depth, and analyses of subjects' verbalizations while studying and summarizing. We consider each of these in turn, and provide some illustrative results.

Summary content. The summary content scores attempt to provide a measure of the information contained in the summary, weighted for importance. To do this, we analyze the text into idea units and form these idea units into a hierarchical structure. The idea units are referred to as propositions, but tend to be larger units than the propositions defined by others (e.g., Fredericksen, 1975; Kintsch & van Dijk, 1978). Essentially, an idea unit corresponds to a simple sentence or to an inferred idea similar in scope to a sentence. The idea units (some directly from the text and some inferred) are organized into a structure with three or more levels: typically we employ the levels of themes, main ideas, important ideas, and unimportant ideas, although in some texts it is not possible to identify a thematic level. These levels correspond, of course, to the different levels of comprehension referred to above. Scores may consist of the number of units at each level, or of a total score, formed by weighting the levels (usually 5 for themes, 3 for main ideas, and 1 for important ideas). The nature of the text determines the number of ideas at each level (the text structure), and the difficulty in agreeing on that structure. The scores that a subject obtains should reflect that subject's cognitive structure for the text information. Content scores are used to score summaries or free recalls of the entire text or can be adapted to score probed recall (short answer) questions.

To illustrate the use of content scores, Figure 5 shows an interaction obtained by Stein and Kirby (1992) in their study of children's learning from text in text absent and present summarization. Summary content scores are significantly and positively related to recall in the absent condition, but not in the present condition. The summary content score appears to be a good index of depth of processing in the former condition, but not in the latter. In the text absent condition, subjects were able to copy directly from the text, so their high content scores do not necessarily reflect greater depth of processing.

Summary depth. Our second approach to measuring depth in a summary relies on rating the complexity of the structure or processes underlying the summary. This is clearly more subjective. In some cases we have felt that it yields a more appropriate index of a summary's depth, although this is not always the case. The depth rating may be most appropriate when students have access to the original text during summarization and copy large amounts of text into their summary—this can yield an unrealistically high content score, although often a disastrously low depth rating.

In our most recent work (e.g., Fuller, 1992), we have used the following scale: 1 point: no coherence; 2 points: limited clusters of coherence; 3 points: linear coherence (following the sequence of the text); 4 points: relational framework, within the bounds of the text; and 5 points: a relational framework that extends beyond the text. These depth ratings focus on the structure and coherence of the summary, and thus resemble Biggs and Collis' (1982) SOLO scores.

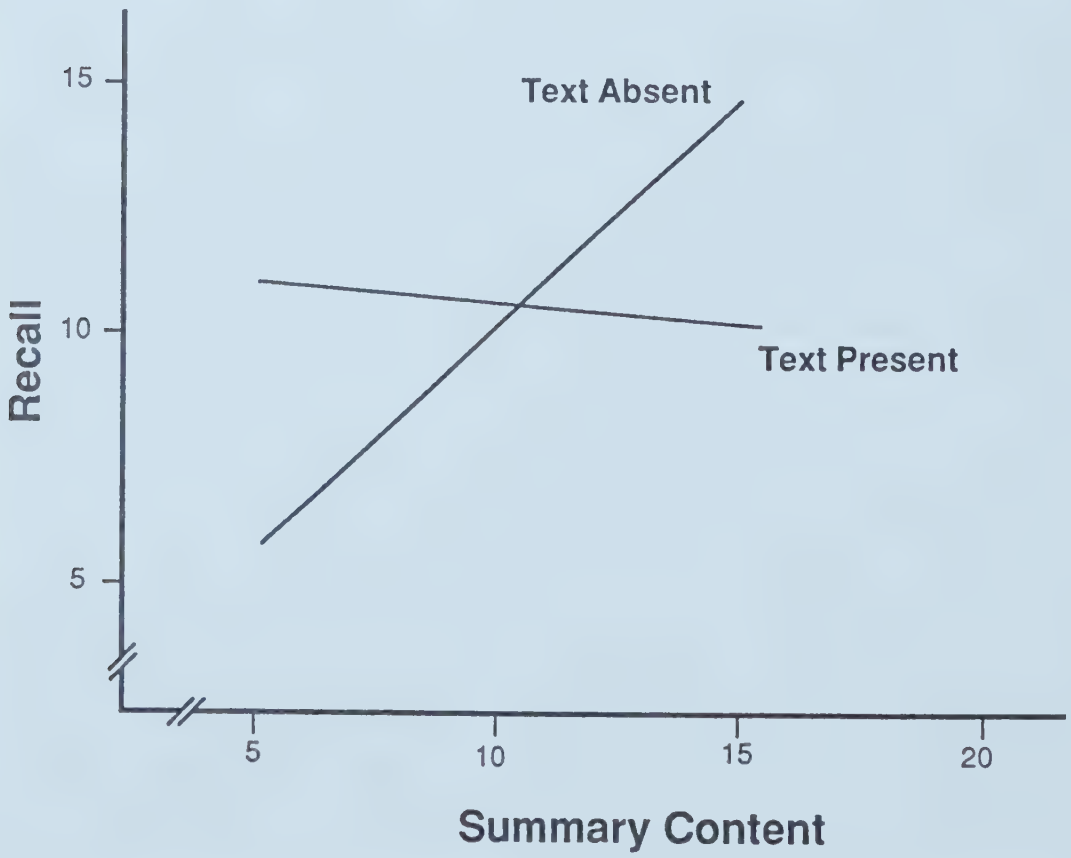


Figure 5. Regression lines showing interaction between task conditions and summary content (adapted from Stein & Kirby, 1992).

Figure 6 shows regression lines obtained in a recent study (Fuller, 1992), relating the depth of processing shown in a summary to recall in two conditions. The conditions differed in the nature of the questions subjects were asked to answer while they studied the text, either relatively surface questions (fact-oriented) or relatively deep questions (oriented toward main ideas, inferences, and interpretations). The deep questions were intended to increase depth of processing. As can be seen in Figure 6, rated depth of processing was not related to recall in the deep questions group; presumably these subjects had already carried out deeper processing during their study, and did not have to do so in their summary. For the surface questions group, however, rated depth of processing accounted for a significant proportion of the variance in recall. Having not been led to depth during study, these subjects needed to attain it in their summaries.

In other work (e.g., Stein & Kirby, 1992), especially with younger readers, we have relied on estimates of the amount of transformation required to go from the text to the summary, as indexed by the number of connectives used, reorderings, lack of copying, and whether or not the writer provided an independent perspective on the text ("What the author was trying to say was ..."). Stein and Kirby found that rated summary depth was significantly related to recall in both text absent and present conditions.

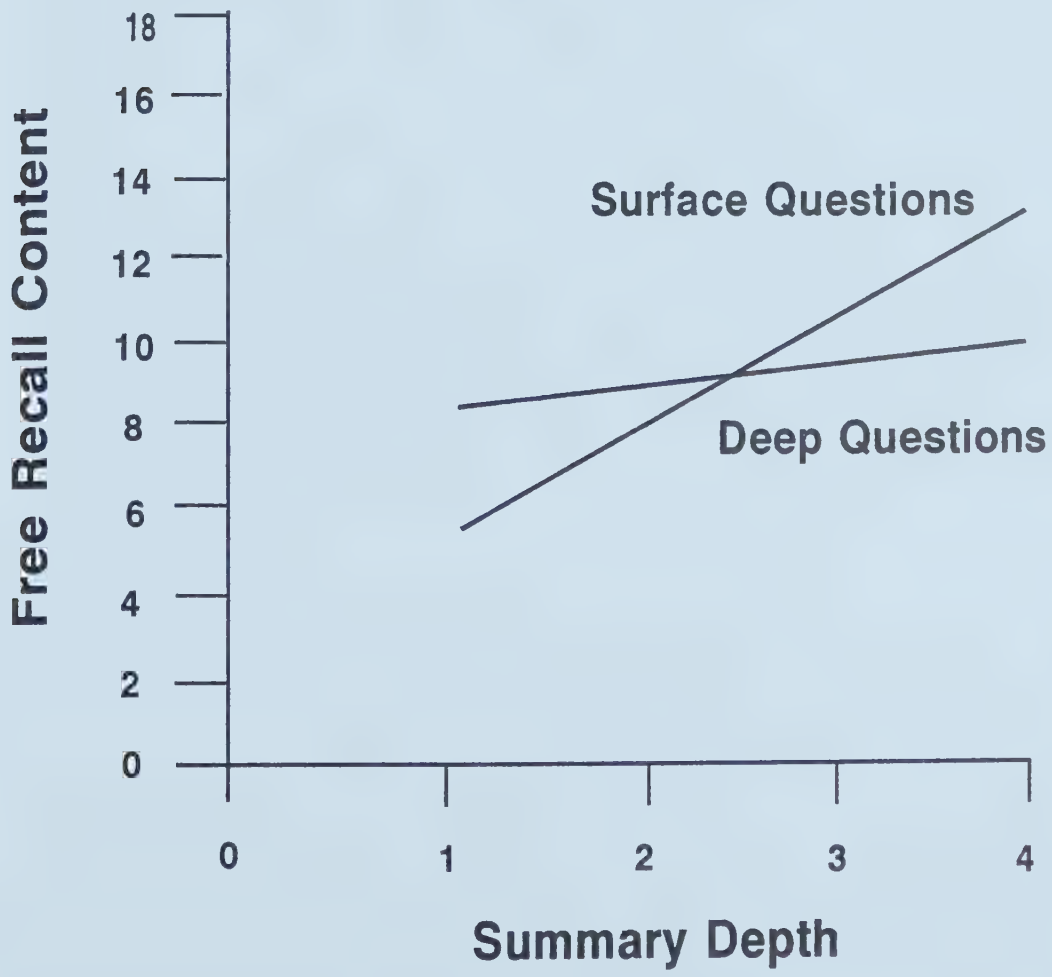


Figure 6. Regression lines showing interaction between task conditions and summary depth (adapted from Fuller, 1992).

Verbalizations during summarization. In the third technique subjects are asked to describe what they are doing as they study and summarize. We developed a 57-category protocol analysis system (Woodhouse & Kirby, 1991), which has subsequently been collapsed into seven or two broader categories (Kirby, Woodhouse, & Ma, 1993). The various categories address different aspects of studying (looking for main ideas, commenting on text difficulty, relating text to prior knowledge, etc.) at different degrees of presumed depth (constructing a main idea as opposed to selecting one from the text). The 57 categories are too cumbersome to employ, but we have had some success with the seven-category version. The two-category version is an attempt to form a surface scale and a deep scale. The seven categories, comprising four surface categories and three deep ones, are shown in Figure 7. The proportion of deep statements made is positively related to performance variables and tends to increase in text absent conditions (e.g., Kirby et al., 1993; Woodhouse & Kirby, 1991).

Depth of Processing During Performance

The final component in Figure 1 concerns learning outcomes. Although we have not studied this explicitly, it is important to remember that subjects exercise options at this point too. For example, a student with a generally deep approach to learning, who has studied the text deeply, may at the point of final performance adopt a more surface or fact-oriented approach. This could be due to performance-related anxiety, to the student's interpretation of the purpose of the test, to fatigue, or to any of a host of other factors that have not been measured. At the moment, all of this is included in the error term. The fact that significant relations between the preceding variables and recall have been shown does not negate the importance of performance factors.

In a sense, most of the study process measures are also measures of performance (a summary is as much a performance as a subsequent free recall). Many of these measures could be applied as successfully to learning outcomes. For example, we also use content scores to assess free recalls; this avoids the difficulty that some subjects copy directly from the text in text present conditions. Biggs and Collis (1982) intended their SOLO Taxonomy as a measure of learning outcomes, not of study processes, which suggests that our depth ratings could also be so used.

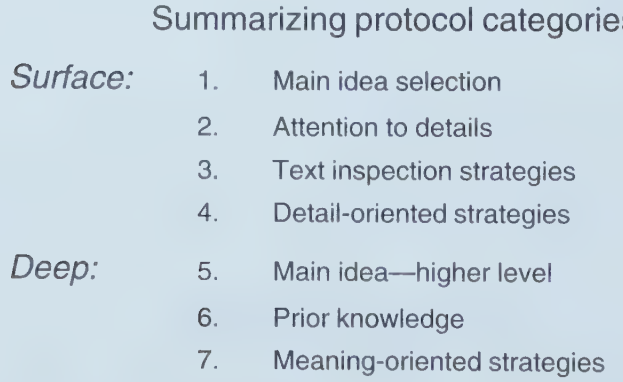


Figure 7. Categories used in protocol analysis.

Implications for Measuring Achievement

How to Do It

It is important to distinguish between what students actually do and what they say they are doing, or even what they say they typically do. What they say they are doing (verbal protocols) may have some value as a research technique, to help validate some measures, but ultimately will be less useful than more objective performance measures. What they say they typically do (their approaches to learning as measured by questionnaire) may have some value in counseling, and may help us understand why some students engage in deeper processing, but it cannot be a substitute for actual performance.

The value of the research we summarize here is that it points to techniques for assessing depth in actual performance, whether during study or on a later test. We see at least five areas of application:

Depth or SOLO ratings. As Biggs and Collis (1982, 1989) suggest, such ratings of quality may be applied to substantial learning products, such as essays or long answers to questions. It is important that the performance be a construction, something generated by the student from memory, rather than something copied from sources or "dumped" from rote memory. Because they attempt to reveal the student's underlying cognitive structure, it is important that the question or prompt not provide too much structure. This condition may be a source of difficulty, because the performance depends critically on the student's interpretation of the goal of the performance. In any particular performance, a student's depth rating could drastically underestimate potential depth. Over time, however, if depth were stressed as a goal, this would become less and less of a problem. It is likely that many students are capable of greater depth than they demonstrate, but do not realize that it is desired (Kirby & Cantwell, 1985; Peters, 1993). Depth or SOLO ratings also offer the potential to provide students with clearer feedback for improving performance.

Content scores. One difficulty with depth ratings is that they are quite global: a five-page essay may yield a single score on a five-point scale. Content scores seem to offer greater potential for guiding students' learning of particular information. Again, it is important that the performance judged be a construction, as copying can yield unrealistically high content scores. There would also be merit in teaching the students text analysis and the content scoring scheme. In this way they could become more aware of text structure and the relative importance of ideas in a text. A text analysis could be used as a concept map to provide students (or to have them generate for themselves) a complete outline of content; these maps could be extended to the thematic level to emphasize that such a level exists and how it is connected to the lower levels.

Short-answer or multiple-choice questions. Although longer, constructed performances are the ideal, it may be useful in some contexts to supplement assessment with shorter, probed recall questions or even multiple-choice questions. Probed recall questions have the disadvantage of providing some structure, so students' performance may overestimate their ability to generate. On the other hand, carefully designed short-answer or multiple-choice questions, based on a text analysis could be helpful. It is important to be clear about what level of information is being assessed (main idea, important supporting idea, unimportant idea) and if an inference is being assessed, what type it is. For

example, inferences may connect ideas at the same level in the text (forming horizontal coherence), or ideas at different levels (forming vertical coherence), or may connect ideas from the text with background knowledge. Most current measures of reading comprehension present a motley collection of items with little regard for level or type.

Dynamic assessment. One difficulty in assessing depth of processing is that some or many students may not know that it is desired and so will not perform as well as they might. One strategy would be to place the students in situations where deeper processing makes more sense or is required. Examples would be text absent summarization and use of poorly structured texts. A second strategy would be to instruct students in the use of deeper processing. By comparing students' performance in standard and depth-encouraging conditions, or pre- and postinstruction, a better measure of ability to engage in deeper processing may be obtained.

Strategy assessment. Underlying all these suggestions is a notion that students' strategies are important, not only in the sense of how they attempt to attain goals, but which goals they seek. Strategies may be inferred from the various depth ratings and content scores, but need to be addressed explicitly in instruction.

Consequences

Although depth of processing is espoused widely as an important goal in education, much of what is done in education works against it. Many students and teachers have conceptions of learning that do not include depth of processing. Many school and classroom practices encourage these conceptions, and defeat attempts for change. At the same time, other notions of teaching and learning are unrealistic, arguing that depth can be attained without content. These notions favor depth, but serve to discredit it as a goal in the eyes of some observers. The effective measurement of depth of processing is an important educational goal, but it needs to develop in conjunction with a greater emphasis on depth of processing in learning and teaching.

Acknowledgments

The research on which this article is based was supported by grants from the Social Sciences and Humanities Research Council of Canada to the first author. The contributions of Jennifer Fuller, Allyson Hadwin, Yamin Ma, Denise Pedwell, Jessica Peters, Jacqueline Simpson, and Bonnie Stein are gratefully acknowledged.

References

Biggs, J.B. (1987). *Student approaches to learning and studying*. Melbourne, Australia: Australian Council for Educational Research.

Biggs, J.B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology,*

Biggs, J.B., & Collis, K.F.(1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J.B., & Collis, K.F. (1989). Towards a model of school-based curriculum development and assessment: Using the SOLO taxonomy. *Australian Journal of Education, 33,* 149-161.

Biggs, J.B., & Kirby, J.R. (1984). Differentiation of learning processes within ability groups. *Educational Psychology, 4,* 21-39.

Biggs, J.B., & Rihn, B. (1984). The effects of intervention on deep and surface approaches to learning. In J.R. Kirby (Ed.), *Cognitive strategies and educational performance*. Orlando, FL: Academic Press.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: Framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Duncan, M.J., & Biddle, B.J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Entwistle, N.J. (1988). Motivational factors in students' approaches to learning. In R.R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 21-51). New York: Plenum Press.

Fredericksen, C. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458.

Fuller, J. A. (1992). *Deep and surface processing in learning from text among baccalaureate nursing students*. Unpublished master's thesis, Queen's University.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research, 56*, 473-493.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*, 161-195.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kirby, J.R. (1988). Style, strategy, and skill in reading. In R.R. Schmeck (Ed.) *Learning styles and learning strategies* (pp. 229-274). New York: Plenum Press.

Kirby, J.R. (1991). Reading to learn: Toward an applied psychology of reading comprehension. In J.B. Biggs (Ed.), *Teaching for learning: The view from cognitive psychology* (pp. 103-125). Melbourne, Australia: Australian Council for Educational Research.

Kirby, J.R., & Cantwell, R.H. (1985). Use of advance organizers to facilitate higher-level text comprehension. *Human Learning, 4*, 159-168.

Kirby, J.R., & Pedwell, D. (1991). Students' approaches to summarization. *Educational Psychology, 11*, 297-307.

Kirby, J.R., Woodhouse, R.A., & Ma, Y. (1993). *Studying in a second language: The experience of Asian students in Canada*. Unpublished manuscript.

Lai, P., & Biggs, J. (1994). Who benefits from mastery learning? *Contemporary Educational Psychology, 19*, 13-23.

Peters, J. (1993). *Summarization strategies, text structure and working memory in recall of text material*. Unpublished master's thesis, Queen's University.

Simpson, J. (1992). *Working memory and the recall of written material*. Unpublished master's thesis, Queen's University.

Stein, B.L., & Kirby, J.R. (1992). The effects of text absent and text present conditions on summarization and recall of text. *Journal of Reading Behavior, 24*, 217-232.

Woodhouse, R.A., & Kirby, J.R. (1991). *A protocol analysis of summarization strategies*. Paper presented to the Canadian Psychological Association, Calgary.

Woodhouse, R.A., Kirby, J.R., Simpson, J., & Hadwin, A.F. (1992, June). *Effects of text structure and summarization processes on recall*. Paper presented at the annual meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Quebec City.

The Alberta Journal of Educational Research *Vo. XL, No. 2, June 1994, 163-175*

Philip Nagy
Ontario Institute for Studies in Education

Cognitive Assessment of Solutions to an Ill-structured Problem Using Forced Classification Analysis

This article illustrates the use of forced classification for assessing solutions to ill-structured problems. The analysis, on discussions by children of a social problem, consists of content categorization, quality ratings within content categories, and analysis by categorical discriminant analysis, with student grade and judged quality as discriminating variables. Results show that FCA can be used to identify ways in which respondents to ill-structured problems can be differentiated using detailed differences in content and quality of responses. The technique is applicable to essays, oral examinations, think-aloud analyses, and other ill-structured problems.

Cet article démontre l'utilisation de la classification forcée pour l'évaluation de solutions à des problèmes sociaux mal structurés. En se basant sur des discussions d'élèves, à propos d'un problème social énoncé, l'analyse consiste à la catégorisation du contenu, au classement de qualité à l'intérieur du contenu de ces catégories, et à l'analyse par catégorie de façon à ce que leur contenu soit analysé. On considérait comme variables discriminatoires le niveau scolaire de l'élève et la qualité jugée par l'élève. Les résultats indiquent que l'analyse par la classification forcée (ACF) peut être utilisée pour identifier différentes façons par lesquelles les répondant(e)s aux problèmes sociaux mal structurés peuvent être différencié(e)s en utilisant les différences détaillées du contenu et la qualité des réponses. On peut se servir de cette technique dans des dissertations, dans des examens oraux, des analyses à haute voix, et dans le contexte de d'autres problèmes mal structurés.

Background

The purpose of this article is to illustrate the use of forced classification analysis (FCA, Nishisato, 1984) for the cognitive assessment of solutions to ill-structured problems. The illustration involves 156 discussions by grades 3 and 6 students, taken in groups of about five. The discussions deal with a family situation. The groups of children are here referred to as subjects. This report focuses on the method of analysis that was used, not on the substance of the problems addressed by the subjects.

The development of the methodology is best understood in terms of concepts drawn from assessment, problem solving, memory research, and the quantification of categorical data. The original study from which these data come (Nagy, 1992) draws also on the literature on the teaching of thinking skills (e.g., Resnick & Klopfer, 1989). This latter body of work is not discussed in any detail.

Philip Nagy teaches in the Measurement and Evaluation program. He is interested in cognitive models of assessment.

Much current measurement research can be viewed as an attempt to generalize technologies originally developed for measurement of simple multiple-choice outcomes to more complex outcomes, such as polytomous items and performance assessments. This report comes at the issue from a different perspective. Consider a continuum of outcome complexity, with responses to multiple-choice items at one end (say the left) and responses to complex, ill-structured problems at the other. The left end of this continuum can be considered the traditional realm of measurement theory, whereas the right end has been the arena for one branch of cognitive science. Recent work in measurement can be viewed as moving from the left end toward the middle, whereas this report should be viewed as moving from the right to the middle.

This is an attempt to introduce quantification to descriptive procedures. As Glaser (1991) states,

> The weak link in the construction of proficiency tests was our descriptions of the performances to be measured; systematic techniques were needed to describe more adequately the components of proficiency at specified levels of competence ... In the present state of the art, the study of performance must be given precedence over or, at the least, equal status with measurement technique for effective approaches to subject-matter assessment. (p. 18)

Assessment and Higher-order Objectives

Writing on the relationship between assessment and the curriculum, many authors (e.g., Nagy, Traub, & MacRury, 1986) have observed that what is most easily assessed tends to become viewed as most important. Some of the more complex goals of education lack appropriate assessment methods (Archbald & Newmann, 1988). Consequently, existing assessment aimed primarily at the less complex goals can be viewed as exercising a narrowing influence on the curriculum. Stiggins (1988), for example, has argued that attempts to examine various approaches to the teaching of thinking skills have been limited by available assessment tools. Despite progress in assessing complex goals (e.g., Baker, Freeman, & Clayton, 1991; Collis & Romberg, 1991; Nickerson, 1989), there is tension between testing and curriculum renewal. Although most would agree that the ability to discuss a complex situation is an important higher-order outcome of schooling, it is not easily assessed and, one might surmise, is not often emphasized in the curriculum.

Problem Solving and Schema Theory

The task used in this study is an ill-structured problem. Frederiksen (1984) summarized the distinction between well-structured and ill-structured problems. Ill-structured problems are more complex, have fewer definite criteria for deciding if a solution has been reached, and lack complete information and a convenient list of accepted procedures. They have higher verbal content and are more context dependent. Most real-life problems would be classified as ill-structured, as would many harder-to-define but important outcomes of schooling, such as the ability to reflect on issues, marshall arguments, and discuss.

Methods for the analysis of complex problems are emerging (Larkin, 1980; Lawrence, 1988; Voss, Greene, Post, & Penner, 1983). Voss and Post (1988) have related this work to schema theory (Anderson, Spiro, & Anderson, 1978), a

system for examining memory for complex events in which schemata act as mental structures or generic scripts. Interpretation of memory data involves matching elements in a specific situation to generic slots or placeholders. Schallert (1982) notes that schemata evolve, becoming more elaborate and specific with experience. Schema theory provides procedures useful for charting complex verbal phenomena in a measurement framework; their use does not imply any direct link with memory research. This allows avoidance of the debate on the reality of schemata (e.g., Abelson, 1981; Alba & Hasher, 1983), while taking advantage of a promising procedure for examining individual differences in verbal performance.

Assessing Quality

A difficulty in applying a schema-based analysis in a measurement context is how to introduce conceptions of quality, or how to decide what constitutes a more complete or sophisticated discussion. Horton and Mills (1984) describe this problem as the lack of an independent definition of depth of processing in the schemata perspective. Discussion of quality can be introduced in at least three somewhat unsatisfactory ways: holistic ratings by judges, informed but arbitrary designation of some categories as better than others, or by imposition of an external criterion measure so that data from some respondents (say older or higher achieving) are defined as better.

A fourth method, used here, begins with categorization of elements of a discussion, yielding a subjects-by-categories matrix indicating the presence or absence (1 or 0) of specific elements in each response. Next, a system for categorizing students' approaches to problems into levels of increasing complexity, adapted from the neo-Piagetian perspective of Biggs and Collis (1982), can be applied to rate the statements by each subject within each category. This replaces the 1s in the data matrix with ratings, thereby introducing an assessment of quality that is nested within a specific content category.

Forced Classification Analysis

With a matrix of subjects by content categories, with entries of either 0 or a rating, the issue becomes how to uncover the categories, and rating levels within categories, that best differentiate among subjects. Discriminant analysis (e.g., Kerlinger & Pedhazur, 1973) is inappropriate because of the nature of the data: the data are typically not normally distributed; the matrix is usually relatively empty; and although the ratings have ordinal properties, it is an empirical question whether "good" solutions can be characterized by a statistic such as the frequency or number of highly rated statements. A method that treats the data as nothing more than unordered categories is preferred.

Forced classification analysis (Nishisato, 1984) provides a more appropriate approach. Based on dual scaling (Nishisato, 1980), the principal components analysis of categorical data, FCA allows a variable to be chosen as the forcing variable and maximally discriminates among subjects across this chosen variable.

An introductory example will clarify the approach. Consider the exemplary data in Table 1. Seven subjects have responded to three questions with three options each. These questions might have a correct answer or merely request an opinion. The initial layout of the data appears in Table 1a. The first step in

Table 1a
Data for Seven Subjects Choosing one of Three Options for Three Questions

Questions		*Q.1*			*Q.2*			*Q.3*		
Options:		1	2	3	4	5	6	7	8	9
Subjects	1	0	1	0	1	0	0	0	0	1
	2	1	0	0	0	1	0	1	0	0
	3	0	0	1	0	0	1	0	1	0
	4	1	0	0	0	1	0	0	1	0
	5	0	1	0	0	0	1	0	0	1
	6	0	0	1	0	0	1	1	0	0
	7	0	1	0	1	0	0	0	0	1

Table 1b

Questions		*Q.1*	*Q.2*	*Q.3*
Subjects	1	$Y_1 X_2$	$Y_1 X_4$	$Y_1 X_9$
	2	$Y_2 X_1$	$Y_2 X_5$	$Y_2 X_7$
	3	$Y_3 X_3$	$Y_3 X_6$	$Y_3 X_8$
	4	$Y_4 X_1$	$Y_4 X_5$	$Y_4 X_8$
	5	$Y_5 X_2$	$Y_5 X_6$	$Y_5 X_9$
	6	$Y_6 X_3$	$Y_6 X_6$	$Y_6 X_7$
	7	$Y_7 X_2$	$Y_7 X_4$	$Y_7 X_9$

dual scaling deletes the 0s, and replaces the 1s with an appropriately subscripted x for the option weight multiplied by an appropriately subscripted y for the subject weight. The result appears in Table 1b.

The second step involves the application of Guttman's principle of internal consistency:

1. Assign as similar scores as possible to subjects choosing the same option, and as different scores as possible for those choosing different options; and
2. assign as similar weights as possible to options chosen by one subject, and as different weights as possible to options chosen by different individuals.

Algorithmically, this is operationalized as follows:

1. Determine nine values of X_j, option weights, to minimize within-subjects differences and maximize between-subjects differences in the values of weights; and
2. determine seven values of Y_i, subject weights, to minimize within-option differences and maximize between-option differences in scores.

This amounts to maximizing the product-moment correlation defined over the x, y pairs in the table.

A forced classification analysis is obtained when one of the "items," a categorical variable such as sex, experience, or grade, is designated as the criterion variable. Option weights are derived to maximize correlation of other variables with the criterion. The process assigns weights to subjects and category ratings to maximize the similarity of subjects who use the same categories (at

Table 2
The Data Array for Forced Classification

Grade	*Categories*																
	11	12	211	212	213	214	22	231	232	233	24	25	26	311	312	32	33
3	2	2	0	0	0	2	3	3	2	0	3	0	2	3	0	1	2
3	0	1	3	0	0	2	2	3	0	0	2	0	0	3	1	0	0
3	0	1	2	0	0	0	2	2	0	1	0	0	0	0	0	0	0
3	2	2	1	0	1	2	2	3	0	2	0	0	0	4	0	0	1
3	2	0	2	0	0	1	2	0	0	1	2	0	0	0	0	0	0
3	0	1	2	0	0	2	2	2	0	0	0	0	0	0	0	0	0
3	0	2	1	0	2	2	2	1	0	2	0	0	0	2	3	0	0
3	2	2	1	2	0	0	2	2	0	0	0	0	0	0	1	1	0
6	0	3	0	0	1	0	3	2	0	0	2	2	0	3	0	2	3
6	0	0	3	0	0	3	3	2	0	0	0	0	0	2	0	0	1
6	0	3	3	0	0	2	1	1	2	0	0	0	0	0	1	1	2
6	0	0	2	0	2	3	2	2	2	1	0	0	0	0	0	0	0
6	0	2	1	0	0	3	2	3	3	1	0	0	0	2	0	0	3
6	3	3	0	0	0	0	2	3	0	2	0	1	0	3	3	3	3
6	1	0	0	0	0	2	2	3	0	0	0	0	0	0	0	0	0
6	0	2	0	0	0	2	2	2	3	0	0	0	0	2	4	2	3

the same rating) and to minimize similarity of those who use different categories. Weights are assigned so that those subjects flagged as being maximally different from each other on the FCA solution can be identified and compared. Table 2 shows a data array from the present study, with the forcing variable, Grade, on the left, and column labels from this study, 1.1, 2.1.1, through 3.3, and ratings in the body of the table (decimals are omitted, and column labels are explained below).

In interpreting the results, the categories that best discriminate among subjects can be identified through their correlations with the solution. Also, the specific rating differences within a discriminating category can be determined. For example, one category might discriminate between those who did not address an issue at all and those who did (i.e., between 0 and any other rating), whereas another might distinguish those with high ratings from those with low ratings. Finally, characteristics of subjects at opposite ends of the solution can be compared on independent variables.

A forcing variable with N levels produces N-1 solutions. Successive solutions account for decreasing proportions of variance, but very small drops in variance occur across these solutions, making them essentially equal in potential importance.

There are some disadvantages to FCA. First, the sampling distributions of analyses are unknown, so no statistical tests are possible. The technique serves as a heuristic device only. Second, subjective decisions are required as to the

number of subjects and categories to characterize and interpret in a given solution. And third, information on the ordering of ratings is lost; the data are treated entirely categorically. Despite these problems, FCA classification offers a potentially useful tool for the final stage of these analyses.

Method

The Sample and Data Collection

The analysis proceeds by creating a system of categories that encompasses all of the ideas expressed by the respondents. Statements are categorized into these slots and then rated according to the following scale, adapted from Biggs and Collis (1982):

1. Makes one point without supporting argument or elaboration;
2. Makes two or three unrelated points without supporting arguments or elaboration;
3. Makes more than two points and relates them to each other with elaboration or synthesizes several ideas with evidence of thought;
4. Makes more than two points and relates them to each other with supporting elaboration, showing depth of thought, well-articulated explanations and original ideas.

The subjects came from grade 3 and 6 classes of nine rural elementary schools located in three school districts in southwestern Ontario. In the original study (Nagy, 1992) the three schools in each district were assigned to one of three methods of teaching thinking skills. This treatment variable is not used in the analysis reported here, but its presence and balanced nature should be noted. Data from two consecutive years were used, so that the intended sample involved 18 classrooms from each grade. One class from each grade was lost due to scheduling difficulties.

Data collection. Students from a class were subdivided into groups of about five and asked to discuss, for five or ten minutes, a problem involving two children, B.J. and Pat. Pat is heavily involved in extracurricular activities, while B.J. is around the house a lot. Both get the same allowance, but B.J. is asked to do more chores than Pat. She or he complains to no avail. The problem, presented in writing, was 140 words; age was not specified, and gender needed to be inferred. In total, 156 such discussions were recorded.

Two variables were used as forcing variables in the FCA analysis: Grade refers to either grade 3 or grade 6; Score refers to independent ratings of the overall quality of each group discussion. This latter variable was derived from ratings by two judges, initially on a 10-point scale, but reduced for this analysis to a 3-point scale. Each judge worked independently, and any disagreements were resolved by discussion.

Matrix Generation

The audiotapes were transcribed and segmented with a change in speaker (Ericsson, 1987). Two analysts prepared a hierarchical category system to encompass the responses made to the problem, using a randomly chosen sample of the data. This process required a few iterations, and the category system underwent minor modifications during application to the full data set.

Next, the analysts rearranged the data so that all statements from a subject coded in the same category were placed together, even if uttered at different

times. Then, they rated these collected statements on the 4-point scale. After a two-month break, the analysts gave holistic scores to the discussions on a 10-point scale, under instructions to aim for equal distribution across the scale. For the categorizations, the ratings, and the holistic scores, the two analysts worked independently and then came together to resolve disagreements. The analysis was done originally using a larger number of categories, which were later amalgamated by assigning to each subject the largest rating within the set of categories so combined. The category system appears as Appendix A; 17 (reduced from 31) categories were used.

Forced Classification

At this point, we have a 17-categories-by-156 subjects matrix, with elements of 0 for no mention of a topic through 4 for a well-thought-out argument. For the forcing variable Grade, one solution is reported, and for Score, two solutions. The 4-point depth-of-processing ratings are called *ratings*. Arbitrary decisions are required for interpretation of the FCA results. The first decision is how many subjects to use in exploring solution dimensions. FCA solutions provide weights to subjects, so that those with high positive weights can be contrasted with those with high negative weights. Forty of the 156 groups, 20 at each end of the solution dimension, were contrasted. This contrast involves examining the rating patterns (on the 0-4 scale) for the two groups. It plays a major role in the second decision.

This second decision concerns how many categories to choose to best distinguish between subjects. FCA provides a correlation of each category with the solution, allowing rank-ordering of the categories for both data sets. Reporting starts with the category correlating highest with the solution. With 20 subjects available at each end of the solution, approximate chi-square tests could be used to decide when the patterns of ratings of subjects at each end of the solution were no longer sufficiently different to report.

Results

Agreement between the two analysts on categorization by content was 91% before resolution; corresponding agreement on depth-of-processing ratings was 90%. For the Score variable, the holistic ratings of overall quality, agreement was 90% on the original 10-point scale, with disagreement of more than one point in fewer than 5% of the cases. These 10-point scores were reduced to the 3-point Score variable for this analysis. Because of the bulk of the data, these reliability figures are based on a sample only, and on the original, more complex category systems.

Variance accounted for in the solutions were for Grade, 80% and for Score, 46% for each of two solutions. Coefficient alphas for all solutions were over 0.96.

The responses of the 156 groups of children were categorized using the 17-category system reported in Appendix A. The bulk of the 1,116 non-zero rated statements were in a few categories: discussion of whether the mother's decision was fair or not (9%), changing chore arrangements (26%), activities (13%), or allowances (19%), and discussing personal experiences (13%). A substantive interpretation of the data is available in Nagy (1992).

Table 3
FCA Solution Using Grade as the Forcing Variable Frequencies on Discriminating Categories for Extreme Groups

Category Number	*Definition*		*Rating* 0	1	2	3	4
1.1	Mother's decision fair	Gr6	13	1	4	2	0
		Gr3	18	2	0	0	0
2.2	Change activities	Gr6	0	1	10	9	0
		Gr3	12	2	3	2	1
2.3.2	Child should pay for activities	Gr6	9	2	3	5	1
		Gr3	20	0	0	0	0
3.1.2	General knowledge or views	Gr6	2	0	8	4	6
		Gr3	10	8	2	0	0
3.2	Seek additional information	Gr6	6	0	12	2	0
		Gr3	18	2	0	0	0
3.31	Group discussion	Gr6	2	3	1	12	2
		Gr3	11	0	8	1	0

[1]This category is unlike the rest, which are content based. It is an assessment of the extent to which the children talked to each other, rather than merely taking turns talking to the moderator.

The single FCA result using Grade (3 versus 6) as the forcing variable is summarized in Table 3, where frequencies of ratings are reported for contrasting groups at opposite ends of each solution dimension. Six categories were identified as distinguishing between the 20 subjects at opposite ends of the solution dimension. There are clear differences, all favoring the grade 6 students, as would be expected. The solution dimension did not achieve perfect separation by Grade, with three grade 6 groups included in the lower scoring group, and one grade 3 in the higher.

On the first FCA solution using Score as the forcing variable (Table 4), eight categories had significantly different distributions across the ratings. Again, the results are striking. This solution makes a clear distinction between high scoring (19 scores of 3 and 1 score of 2) and low scoring (19 scores of 1 and 1 score of 2) groups. FCA has identified 40 of the 156 groups that differ in consistent ways across eight of the 17 categories. These eight categories include five of six from Table 3, the Grade results, plus three additional ones.

The second FCA solution using Score as the forcing variable is quite different. It makes a slightly less clear distinction between medium scoring discussions (one score of 1, 16 of 2 and 3 of 3) and a mixture of high and low scoring ones (10 scores of 1, 1 of 2, and 9 of 3). Six categories exhibited significantly different ratings patterns, but unlike the first two reported solutions, these patterns are not all in the same direction. The data for three categories, 2.1.1, 2.1.4, and to some extent 2.3.3, favor the medium scoring groups, and Category 2.4 favors the mixture of low and high scoring. The remaining two categories, 2.3.1 and 3.1.2, exhibit unusual patterns contrasting bunched ratings with fairly even distributions across all ratings. This solution seems to have identified unique ways in which middle quality discussions differ from each of high quality and poor quality ones.

Table 4
FCA Solution Using Score as the Forcing Variable
Frequencies on Discriminating Categories for Extreme Groups

Category Number	*Definition*		*Rating* 0	1	2	3	4
First Solution							
2.1.1	BJ do less or Pat do more	HI	5	1	9	4	1
		LO	13	5	2	0	0
2.1.4	Schedule or share chores	HI	2	1	5	8	4
		LO	17	0	3	0	0
2.2	Change activities	HI	1	0	8	11	0
		LO	4	7	9	0	0
2.3.1	Gear privileges to chores	HI	0	0	5	12	3
		LO	10	7	3	0	0
2.3.2	Child should pay for activities	HI	8	1	4	6	1
		LO	19	1	0	0	0
3.1.2	General knowledge or views	HI	0	0	7	5	8
		LO	14	6	0	0	0
3.2	Seek additional information	HI	8	1	9	2	0
		LO	18	2	0	0	0
3.3	Group discussion	HI	3	1	6	8	2
		LO	19	1	0	0	0
Second Solution							
2.1.1	BJ do less or Pat do more	ME	2	4	10	4	0
		HL	11	4	3	1	1
2.1.4	Schedule or share chores	ME	0	2	10	8	0
		HL	13	3	2	2	0
2.3.1	Gear privileges to chores	ME	0	1	19	0	0
		HL	5	5	3	5	2
2.3.3	Keep allowances equal	ME	12	0	7	1	0
		HL	12	4	4	0	0
2.4	BJ talk to friend, family	ME	15	5	0	0	0
		HL	5	2	6	7	0
3.1.2	Role of experience	ME	5	5	10	0	0
		HL	11	2	2	2	3

HI, ME, LO; high, medium or low; HL; a mixture of high and low

Summary and Discussion

The purpose of this article is to explore the capabilities of FCA for the cognitive analysis of ill-structured problems. One major advantage of this procedure is the high degree of reliability attained in the subjective decisions. By separating decisions on content from those on quality, and by constraining the quality decisions, levels of agreement far surpassing those in typical holistic grading (e.g., Nagy, Evans, & Robinson, 1988) were achieved.

Results forcing on Grade showed clear differences. However, the results for the variable Score were of much greater interest, for two reasons: first, the discriminating categories and identified subjects in the first solution were different from those based on Grade, indicating that the two bases for distin-

guishing were not identical; and second, the second solution found patterns of contrasting strength, rather than showing a consistent advantage for one group over the other, as in the Grade results.

It is worth commenting on the numbers of categories used in the analyses. As mentioned above, the analysis began with a much more complex category system of 31 rather than the present 17 categories. Earlier analyses with the larger category systems did not produce such clear results, suggesting that an overly complex category system can make distinctions better left unmade for the purposes of forced classification analysis.

FCA appears useful when used with two- and especially three-level forcing variables. When a three-level variable like Score is used, FCA appears to be very powerful; it can find interesting contrasts and can show different bases for making different distinctions (i.e., medium from high and low) on the forcing variable.

We have relied on methodology taken from cognitive research, and it is reasonable to examine the relationship between this work and schema theory. Because this it group data, any direct connection with memory per se is problematic. Voss and Post (1988) link methodology to theoretical framework. The intent of this work is development of assessment devices for educators. Thus the theoretical concern is to represent the possible approaches to a problem and to find those aspects that reflect differences in quality. When faced with the choice between richness of detail and operational simplification, a concern with operational measurement has favored the latter. Cognitive scientists seeking in-depth meaning in rich detail may object to this adaptation of one of their principal modelling devices. However, from the perspective of a measurement theory attempting to cope with increasing complexity in outcome data it is useful.

Although this work does not advance schema theory, it does advance analysis of ill-structured problems. Lawrence (1988) examined magistrates' thinking, using three magistrates dealing with three cases. Voss et al. (1983) investigated the lack of productivity of the Soviet agricultural system using fewer than 20 subjects. Relying heavily on this earlier work, the present study has moved to considerably larger sample sizes and some quantification while maintaining what appear to be comparable levels of detail and research costs.

How has measurement profited by this exercise? What has been accomplished by this complex analysis that might not have been done more traditionally, by developing a marking scheme or training judges in holistic scoring? There is ample evidence that teachers require assistance in assessing higher-order outcomes of instruction (Haertel, 1986; Stiggins, 1988). Neither the identification of what constitutes higher-order thinking nor the development of appropriate marking systems is a trivial task. Both require a theoretical framework. Further, there are calls from those studying the assessment practice of teachers (e.g., Stiggins & Bridgeford, 1985) for more focused methods. Systematic research on the bases teachers use for grading is thin (Stiggins, Frisbie, & Griswold, 1989). A method allowing comparison of practice with a well-developed image of what constitutes the target qualifies as a focused method, both for research and for the improvement of practice.

The analysis has proven useful in dissecting complexity in a systematic manner. It has provided a "territorial map" of a complex domain that might be expanded by use of similar data from responses to carefully constructed scenarios. Such work is made easier and less expensive through the use of such a tool. Beyond this particular data set, the technique provides a general method for examining ill-structured problem solving, which plays a prominent role in many curricula, particularly the social sciences.

On the side of caution, forced classification lays open possible structure in the data, but does not replace careful exploration by other means. As with any matrix decomposition method, peculiarities in the data can lead to uninterpretable results. The final test of forced clasification is whether the results make sense in terms of the raw data.

The need for additional careful examination does not diminish the value of the analysis. The method has done more than find two unusual sets of subjects, along with an idiosyncratic way in which they differ; it has identified and exaggerated real differences that exist in the entire sample. If cross-tabulations of grade by rating level are examined for all 17 categories in the data (a tedious process), it is possible to validate the results of Table 3. The cross-tabulations for all 156 groups mimic the data for the 40 groups in Table 3, although the proportions are somewhat more attenuated. Forced classification is efficient in finding differences, but benefits beyond mere time saving are evident when the forcing variable has more than one solution, such as Score. Again, the full set of 17 cross-tabulations reveals no evidence disconfirming the patterns reported in Table 4. However, although these results may be validated by examination of cross-tabulations, they could never be discovered by such examination.

The ability to discuss an ill-structured problem is one example of an important class of somewhat neglected outcomes of schooling. Other members of this class include essays or oral presentations where an opinion must be defended. How "ill" the structure is varies: we can defend marking schemes for explanations of the causes of historical events better than we can views on current ethical or social dilemmas. This report is intended to spark interest in a meeting of cognitive and measurement perspectives on an important issue.

Acknowledgment

The study on which this report is based was funded by the Ontario Ministry of Education.

References

Abelson, R.P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715-729.

Alba, J.W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203-231.

Anderson, R.C., Spiro, R.J., & Anderson, M.C. (1978). Schemata as scaffolding for the representation of information in connected discourse. *American Educational Research Journal, 15,* 433-439.

Archbald, D.A., & Newmann, F.M. (1988) *Beyond standardized testing.* Reston VA: National Association of Secondary School Principals.

Baker, E.L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing.In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Toronto: Prentice-Hall.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning.* New York: Academic Press.

Collis, K., & Romberg, T.A. (1991). Assessment of mathematical performance: An analysis of open-ended test items. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition.* Toronto: Prentice-Hall.

Ericsson, K.A. (1987). Theoretical implications from protocol analysis on testing and measurement. In R.R. Ronning, J.C. Conoley, J.A. Glover, & J.C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 191-226). Hillsdale, NJ: Erlbaum.

Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research, 54*, 363-407.

Glaser, R. (1991). Expertise and assessment. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition* (pp. 17-20). Toronto: Prentice-Hall.

Haertel, E. (1986). *Choosing and using classroom tests: Teachers' perspectives on assessment.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Horton, D.L., & Mills, C.B. (1984). Human learning and memory. *Annual Review of Psychology, 35*, 361-394.

Kerlinger, F.N. & Pedhazur, E.J. (1973). *Multiple regression in behavioral research.* Toronto: Holt, Rinehart & Winston.

Larkin, J.H. (1980). Teaching problem solving in physics: The psychological laboratory and the practical classroom. In D.T. Tuma & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research* (pp. 111-125). Hillsdale, NJ: Erlbaum.

Lawrence, J.A. (1988). Expertise on the bench: Modelling magistrates' judicial decision-making. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. 229-259). Hillsdale, NJ: Erlbaum.

Nagy, P. (1992). The assessment of group discussions and complex problem solving: Potential contributions of schema theory. In D.J. Bateson (Ed.), *Classroom testing in Canada: Proceedings of the second Canadian conference on classroom testing* (pp. 20-37). Vancouver, BC: University of British Columbia, Centre for Applied Studies in Evaluation.

Nagy, P., Evans, P., & Robinson, F. (1988). Exploratory analysis of disagreement among holistic essay scorers. *Alberta Journal of Educational Research, 34*, 355-374.

Nagy, P., Traub R., & MacRury, K. (1986). *Strategies for evaluating the impact of province-wide testing.* Toronto: Ministry of Education.

Nickerson, R.S. (1989). *Educational Researcher, 18*(9) (Special Issue).

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1984). Forced classification: A simple application of a quantification method. *Psychometrika, 49*, 25-36.

Resnick, L., & Klopfer, L. (Eds.). (1989). *Toward the thinking curriculum: Current cognitive research. ASCD yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Schallert, D.L. (1982). The significance of knowledge: A synthesis of research related to schema theory. In W. Otto & S. White (Eds.), *Reading expository material* (pp. 13-48). Toronto: Academic Press.

Stiggins, R.J. (1988). *The nature and quality of teacher-developed classroom assessments.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Stiggins, R.J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286.

Stiggins, R.J., Frisbie, D.A., & Griswold, P.A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement Issues and Practice, 8*(2), 5-14.

Voss, J.F., Greene, T.R., Post, T.A., & Penner, B.C. (1983). Problem solving skill in the social sciences. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research theory* (pp. 165-213). New York: Academic Press.

Voss, J.F., & Post, T.A. (1988). On the solving of ill-structured problems. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. 261-285). Hillsdale, NJ: Erlbaum.

Appendix A
Category System for Group Responses

Category Number	*Definition*
1.	Diagnosis
1.1	Mother's decision fair or neutral
1.2	Mother's decision unfair
2.	Solution
2.1	Change chores
2.1.1	BJ do less or Pat do more
2.1.2	Leave as is
2.1.3	Get additional help
2.1.4	Schedule or share chores, co-operate
2.2	Change activities
2.3	Change allowance
2.3.1	Gear allowance, privileges to chores
2.3.2	Children should pay for activities
2.3.3	Stay the same, both increase, decrease
2.4	Communicate: BJ talk to friend, family
2.5	BJ get a job
2.6	Negative behaviour
3.	Problem solving
3.1	Reflections of experience/knowledge
3.1.1	Specific personal experience
3.1.2	General knowledge or views
3.2	Seek additional information
3.3	Group discussion

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 177-193*

Philip H. Winne

Lorene Gupta

and

John C. Nesbit
Simon Fraser University

Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics

Computer-based adaptive learning systems that afford a high degree of learner control in studying are an excellent medium for collecting extensive data on individual differences. STUDY (Winne, Jones, & Field, 1992) is such a system for writing and delivering adaptive tutorials. As learners engage in a tutorial with STUDY, navigating through content and applying studying actions such as underlining, taking notes, requesting elaborations, and attempting test items, STUDY creates detailed time-stamped records of the learner's interactions in a log file. We are developing a methodology based on the mathematics of directed graphs for analyzing log file data. In this approach a sequence of study actions is reduced to a set of nodes representing action types and a set of links representing a temporal relation called followed-by. *This article describes elementary graph theory statistics and features of graphs that can characterize individual differences in cognitive strategies used in studying.*

Les systèmes informatisés d'apprentissage adaptif qui permettent un degré élevé de contrôle sur l'enseignement personnalisé en fonction de l'apprenant(e) sont un moyen excellent de receuillir plusieurs données sur les différences individuelles. STUDY (Winnie, Jones & Field, 1992) est un tel système d'ordinateur qui permet la composition et la présentation de modules d'études adaptifs sur ordinateur. Quand les apprenant(e)s s'engagent à travailler avec les modules de STUDY, ils doivent naviguer à travers le contenu et faire l'application de différentes méthodes d'étude telles que souligner, prendre des notes, solliciter des élaborations, et essayer de répondre aux questions d'examen. STUDY enregistre de façon chronométrée à l'aide d'un totalisateur intégrée à même le logiciel les interactions des apprenant(e)s dans un fichier de receuils de données. Nous sommes en train de développer une méthodologie pour l'étude de ces receuils de données selon la signification mathématique contenu à l'intérieur des graphiques indiqués par le système dans les modules. Dans cette approche, une série de différentes méthodes d'étude peut être réduite à un groupe de modules qui représente différents types de moyens ainsi qu'un groupe de liens qui représente la relation temporelle que l'on appelle suivi-par. Cet article décrit comment la théorie des statistiques des graphiques élémentaires et les différents aspects de ces graphiques peuvent bien caracté-

Philip Winne is a professor of education and psychology and an instructional psychologist specializing in theories of complex academic learning and adaptive learning systems that promote learning and motivation.
Lorene Gupta is a doctoral candidate in the Joint MA/PhD program in psychology and education. She is studying applications of computer technologies that promote students' active cognitive engagement in learning.
John Nesbit received his doctorate from the University of Alberta in 1988 and is presently a research associate in Project STUDY. His research focuses on the psychology of computer-based instruction.

riser les différences individuelles dans le contexte des stratégies cognitives qu'utilisent les apprenant(e)s lorsqu'ils ou lorsqu'elles étudient.

As students study, they engage in a stream of varied and interacting cognitive activities: selecting information, seeking information, rehearsing, elaborating, self-testing, and so forth. Some of these cognitive events can be traced (Winne, 1982, 1992) by observable actions the learner performs during studying such as underlining, looking up a definition, copying information into a notebook, circling a feature in a diagram and drawing a line to link it to a portion of a text, and attempting to answer a test item. Interspersed among these cognitive and behavioral events is metacognition, strategies the learner engages to guide selecting and applying each studying action, and adapting tactical sets of those acts. At yet another level, there is a dynamic interplay among (a) these cognitive events; (b) more persistent individual difference variables such as prior knowledge and perceptions of self-efficacy; and (c) instructional interventions such as adjunct questions and feedback (Corno & Snow, 1986; Winne, 1993a, 1993b). At this level, questions address how students self regulate (Zimmerman & Schunk, 1989).

A methodology that could selectively examine a stream of observed activities at theoretically different levels—acts, tactics, strategies, and self-regulation—could contribute significantly to advancing instructional theory and educational practice. Contemporary empirical methods suffer shortcomings in representing temporally unfolding cognitive engagements that constitute a learner's expressions of knowledge, motivation, cognition, metacognition, and self-regulation. Classical Fisherian experimental designs, for example, aggregate data that reflect theoretically separate cognitive events constitutive of individual differences. As well, this methodological approach does not examine patterns in time that reflect using a strategy. Retrospective questionnaires and stimulated recall protocols about cognition engaged in an earlier setting (Eriksson & Simon, 1993) have potential to represent such patterns, but because they do not gather data as cognition happens, there is the possibility that learners misrepresent cognition and motivation by engaging in reconstructive processes. Think-aloud protocols (Eriksson & Simon, 1993) may eliminate misrepresentations attributable to reconstructive recall, but they risk interfering with the very individual differences a researcher seeks to study (Genest & Turk, 1981).

We are attempting to develop methodologies for research into instruction and individual differences that reduce these shortcomings and, particularly, that reveal patterns of cognition latent in sequential records of a learner's actions while studying. Our instrument for gathering data, briefly described in the next section, is a computer-based adaptive learning environment called STUDY (Winne, Jones, Field, & Nesbit, n.d.). STUDY is a milieu where a student studies and in the course of that activity generates and traces cognitive and metacognitive events. Subsequently, we introduce analytical techniques we are exploring for researching cognitive and metacognitive individual differences in students' studying behavior.

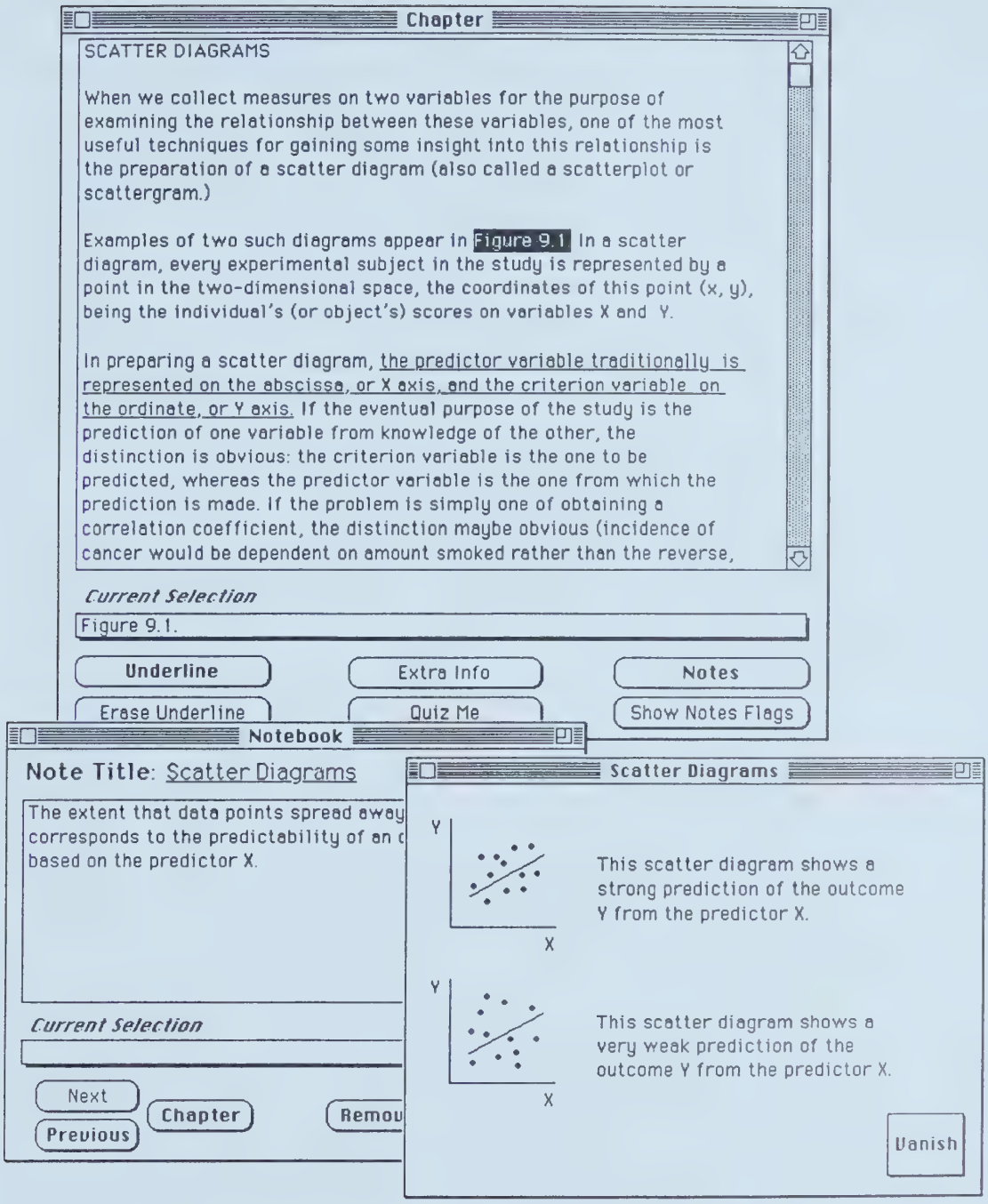


Figure 1. Screen shot of STUDY at an early point in a tutorial.

STUDY: A Tool for Instructional Research into Cognitive Tactics and Strategies

STUDY is a computer system for writing and delivering adaptive learning tutorials on Macintosh™ computers. The system is an amalgam of HyperCard™, which provides the user interface, Smart Elements™, an object-oriented expert system used to represent information and to control the behavior of the tutorial, and routines written in C that glue together the two subsystems and support authoring operations. In STUDY's authoring environment, a designer creates a tutorial by specifying windows, buttons, and fields of text (see Figure 1) and writing rules that define how STUDY interacts with the learner. In a typical tutorial a student would study by viewing text and graphics, clicking on text fields and buttons to display windows containing

ancillary information, attempting test items, and receiving feedback. One goal of the program of research of which STUDY is a part is to investigate individual differences in cognitive strategies that learners use, develop, and can be taught.

STUDY is an adaptive system in that during a tutorial data representing a learner's behavior can be operated on by rules that tailor instructional events to the learner's prior and evolving knowledge, motivation, and strategies. For instance, the designer may identify a continuous series of characters in a text as a concept and program a rule so these words or phrases function as hypertext buttons. During a tutorial, if a student clicks on a concept, a rule may fire that displays a window asking the student to select a type of information (e.g., definition, example, linked concepts) to be displayed by STUDY. The rule also can contain conditions so that another student with a different log of data who clicked on that concept would be shunted directly to an introductory explanation and an accompanying test item about prerequisites.

Following is a partial inventory of learner actions (see Winne, 1992; Winne, Jones, & Field, 1992) used later in this article, which have been researched in instructional psychology: underlining (Peterson, 1992), notetaking (Kierwa et al., 1991), feedback (Butler, Winne, & McGinn, n.d.), viewing examples (Walczyk & Hall, 1989), testing (Glover, 1989), and making attributions (Graham, 1991). During a tutorial, STUDY writes to a log file time-stamped records of all student actions (excluding window scrolling) and all significant actions performed by STUDY (e.g., firing rules, opening windows). By carefully designing tools for the learner to use during a tutorial so that their use validly represents cognitive activities of theoretical interest (Winne, 1982; 1993a), data in the log trace how the learner studies, what the learner achieves, and the conditions on which the learner exercises tactical, strategic, and self-regulatory actions.

Learner Action	*STUDY or Learner Behavior*
Select Text	A portion of text being studied may be selected by a click-and-drag operation. This action is usually a prerequisite for using the text in a subsequent action.
Underline Text	Selected text is underlined. The underline is maintained whenever the text is later viewed by the learner.
Classify Text	The learner or STUDY classifies selected text, for example, as a main idea, a detail, an example or elaboration, or a definition.
Make Note	A window opens in which the learner can enter text. Information entered in a note window is pinned to text the learner has selected in the chapter text or, if no text is selected, to the chapter as a whole. STUDY asks the learner to classify the note as a question posed, a paraphrase of the selected text, a summary, or links made between information.
Request Elaboration	A window opens to present information related to the selected text. The elaboration may be in the form of a

	definition, an analogy, a diagram, an explanation, or a review.
Request Test	An item testing current material is administered. The learner may specify whether the item tests recall or inference.
Request Feedback	After responding to a test item, the learner may choose to receive no feedback, right/wrong feedback, or explanation feedback.
Attribute Performance	The learner specifies an attribution to ability, effort, luck, ease/difficulty of item, or use of a studying strategy.
Calibrate Understanding	The learner estimates comprehension of currently selected sections of text.

An Introduction to Analyzing STUDY's Log Files

In our terms a log file is a chronological record of events occurring in one tutorial with one student. In processing a log file, the analyst filters out items that have no relevance to a particular analysis, rewrites the items of interest to a new file, classifies each as belonging to one of several action types, and denotes each action type by a unique symbol. The resulting temporally ordered string of symbols is called an action sequence. The grand action sequence may be subdivided into smaller action sequences if desired. Action sequences are the primary units of analysis used in describing a student's studying activities.

The most direct measures for describing an action sequence are based on counts of actions such as the number of times a learner retreated in studying a text (e.g., Goldman & Saul, 1990) or the proportion of studied idea units underlined (Johnson, 1988). Useful though such descriptions may be in describing individual differences in the prevalence of comprehension monitoring, counts of action types and statistics derived from them (proportions, means, variances, and so forth) ignore temporal and other relations among multiple actions. We strive to capitalize on such temporal relations. Our hypotheses are that temporally unfolding actions create tactical and strategic structures, and that such structures may be inter- and intraindividual differences that add important information to theoretical accounts about how achievements are created in the course of studying (Winne, 1993a; Winne & Butler, 1994).

In this article we explore analytical tools, based on graph theory, that focus on transitions between actions. In this regard our approach to studying learner-system interactions parallels recent and independent work by Guzdial (1993) who has used Markov chains to analyze programming strategies adopted by science students. Following, we introduce descriptors or *statistics* in elementary directed graph theory and consider how these statistics might reflect individual differences of interest to a cognitive theorist.

An Adjacency Relation and Individual Differences

Consider transitions between two actions, *a* and *b*, observed within a longer sequence of actions generated in a tutorial with STUDY. Suppose *a* is underlining and *b* is making a note that relates *a*'s information to the student's prior knowledge. Transitions among events constitute a relation we call *followed-by*.

If *a* is always followed by *b*, and *b* is always preceded by *a*, the student has composed (Anderson, 1983) a two-action tactic, *a*→*b*, for studying. If we assume the student is a goal-directed, active cognitive agent (Winne & Butler, 1994), each tactic serves a particular goal. Different goals are approached by different actions, tactics, or strategies. Observing a tactic such as *a*→*b* marks several kinds of information about individual differences. First, the student (a) has conditional knowledge about states of cognitive engagement during studying that can trigger the tactic *a*→*b*. On observing multiple uses of the *a*→*b* tactic within a longer sequence of other actions, we also infer that the student (b) attended to cognitive states, (c) discriminated some cognitive states from others, and (d) adapted behavior to accommodate particular states.

We hold that conditional knowledge, attention, discriminated action (adaptation), and the goals that learners seek are core elements that constitute individual differences. If a second student frequently uses *a* separately from *b* or replaces *b* with a sequence *c*→*d*→*e*, this variance in cognitive engagement reflects an individual difference we are interested in studying. Depending on other information about the students and their success at the common task they were assigned, the first student's composition of the three-step sequence of procedural knowledge *c*→*d*→*e* into the unit *b*, and that student's automatization of *a*→b may characterize expertise. Or it may reflect a hobbling mechanization of thought reminiscent of classic findings about negative transfer in problem solving (Luchins & Luchins, 1959). Parallel reasoning applied to one student's intraindividual variation across situations would view such variance as reflecting cognitive development.

Actions

Creating an adjacency matrix is the first step in an analysis using the followed-by relation. Here is a simple example using a small subset of actions available in STUDY and a short action sequence:

E = Ask for an elaboration of selected text.
N = Make a note.
U = Underline a portion of text.
Sequence A: U U N U U U E N U N

Construct an N x N matrix where N is the number of possible actions (three in this example). Label rows and columns with the action symbols. Here we arbitrarily use alphabetical order: E, N, U. In cell (i, j) tally all occasions when action i is followed by action j immediately. Thus the first tally representing the first transition U→U is tallied in cell (3, 3). The next tally, representing the transition from the second U to the first N, is tallied in cell (3, 2). The result,

Table 1
A Weighted Adjacency Matrix and an Adjacency Matrix

	E	N	U		E	N	U
E	0	1	0	E	0	1	0
N	0	0	2	N	0	0	1
U	1	2	3	U	1	1	1

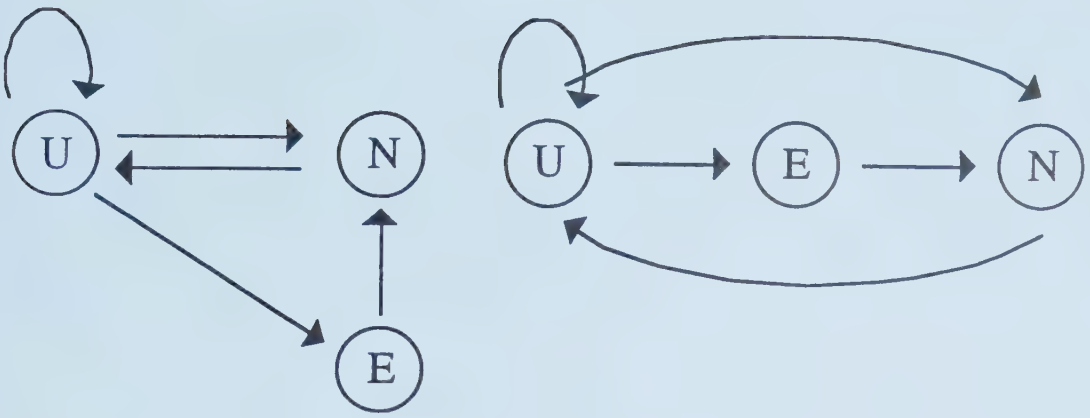


Figure 2. Graph of sequence A showing how repositioning nodes alters the impression conveyed.

illustrated in Table 1, is a weighted adjacency matrix showing the frequencies of all action pairs observed in Sequence A. A weighted adjacency matrix can be simplified to show merely whether action i was followed by action j at least once in the sequence by replacing each non-zero entry in the weighted adjacency matrix with 1, producing an *adjacency matrix*.

Graphing the Followed-by Relation

A graph represents information in terms of a finite set of elements and relationships among the elements. In the example, the elements are actions E, N, U, and there is one relation, followed-by. The left panel of Figure 2 graphically represents the adjacency matrix in Table 1. In Figure 2, each action is represented as a labeled node. For each 1 in row i, column j of the adjacency matrix, a directional link is drawn from the node representing action i to the node representing action j. A self-loop is shown for action U. The graph is *directed* because links represent the directional relation, followed-by. A graph, like Figure 2, is *connected* if none of its nodes is isolated from other nodes. A *weighted graph* labels links using weights in the weighted adjacency matrix.

A directed graph can be visually deceptive because an adjacency matrix implies nothing about where to position nodes. Alternate spatial arrangements may convey different impressions. For example, the right panel of Figure 2 suggests a more "linear" approach to studying.

A Simulated Data Log

To illustrate graph theoretic analyses of action sequences, we generated a log representative of a real STUDY tutorial.[1] One of us (Gupta) kept a record of studying actions while reading the first five sections of Chapter 9 in *Statistical Methods for Psychology* (Howell, 1987) in a course taught by another of us (Winne). Gupta used the three actions (U, E, N) already illustrated, and two more. Although a book can not offer test questions and feedback, as STUDY can, Gupta self-generated test questions (T) and searched the text for feedback about her answers to her self-generated questions (F). T→F is a metacognitive tactic she uses to enhance comprehension (Haller, Child, & Walberg, 1988).

Figure 3 displays the weighted graph, which we call the ALL graph, representing the action sequence across all five sections of the text. In the ALL graph, we excluded the four followed-by links between actions that span sections of the chapter. This reflects our assumption that sections are major

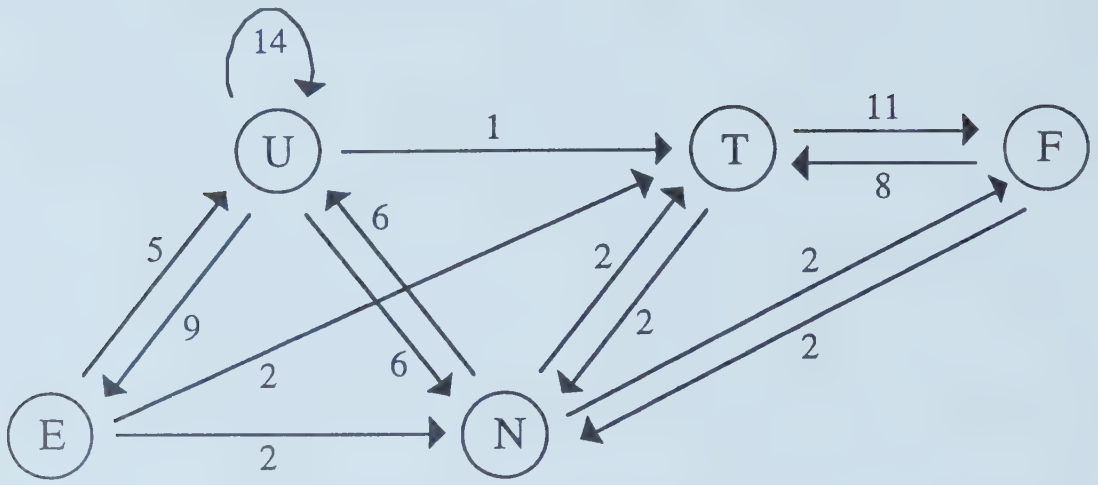


Figure 3. ALL graph of studying actions (transitions between sections are excluded).

content boundaries that probably bound studying strategies and tactics as well. The data corroborate our assumption: the transition F→U never occurred within a section but was observed at most section boundaries.

Graph Concepts and Statistics for Making Interpretations about Strategic Learning

Is a Student Strategic?

Contemporary instructional theories assume that students engaging in learning are goal-directed agents. If this is true, patterns in log files of students' studying should be found that could be interpreted as tracing preferences for or conditional knowledge about actions, tactics, and strategies. For example, in an interview after she studied the text, Gupta indicated her presumption that access to feedback was in one-to-one correspondence to a specific self-generated test question, a pattern like the rigid $a \rightarrow b$ tactic we used to introduce the followed-by relation. Empirically, Gupta was not as consistent as she believed herself to be. The ALL graph shows that she twice made a note preceding feedback and twice made a note following a (self-generated) test item. In these instances, we can infer that she adapted her general tactic to specific, local cognitive conditions.

Gupta's studying operationalized other patterns that reduced the theoretical maximum of N^2 pairwise links within the 5-tuple {U, N, E, T, F}. After examining the log file and interviewing Gupta about her studying and her understanding of STUDY's design capabilities, it was clear that she deliberately excluded links T→U and T→E, as well as self-loops E→E and N→N. Gupta did not know that STUDY could be designed to permit students to return to the text or ask for elaboration immediately after a test item is posed. Nor did she realize that STUDY can be designed to allow students to make two successive but unique elaborations or two successive but unique notes. Gupta also presumed that STUDY would never provide two successive but unique feedback messages, F→F, say, one about content and another about motivation. These tacit rules express individual differences. They reduced the number of theoretically possible followed-by links from 25 to 20, constraining Gupta's ALL graph to a maximum of 80% of STUDY's canonical graph.[2]

The degree to which constraint has been imposed on a canonical adjacency matrix can be measured by two statistics adapted from graph theory, density and cohesion. We interpret these statistics to reflect preference or conditional knowledge.

Density is a graph theoretic statistic, ranging between 0 and 1, that compares the number of links observed in a graph to the number of possible links if all nodes were mutually linked. For N nodes, the number of possible links, including self links (loops), is N^2.

$$\text{density} = \frac{\text{\# links in graph}}{N^2}, \qquad (1)$$

where N is the number of nodes in the graph.

Based on Gupta's reduced canonical graph, we replaced N^2 in equation 1 with 20 instead of 25. With this change, the density of Gupta's ALL graph is .70 relative to her canonical graph for the 5-tuple {U, N, E, T, F}. By reflecting density (reversing its scale by subtracting observed density from the maximum of 1.00), we propose a statistic K_d that measures the facet of individual differences we referred to earlier as discriminated action. K_d is .30 for Gupta's ALL graph relative to the canonical graph. Her cognitive activities, reflected through studying actions that trace her cognition, were constrained by her rules. We might ask the same question of Gupta's studying each of the text's five sections. Table 2 presents K_d values for each section's graph of studying normalized relative to the canonical ALL graph. When observed sequences have sufficient length (sample size), variations in density across sections indicate intraindividual differences in discriminated action when studying different sections of a text.

A second measure of discriminated action that a student exercises during studying is given by the graph theoretic statistic called cohesion. Cohesion measures the proportion of links in a graph, excluding self-loops, that are two-way; that is, for nodes A and B, there is a link from A to B and a second link from B to A. Cohesion is defined in equation 2.

$$\text{cohesion} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} \cdot x_{ji}}{\frac{N^2 - N}{2}}, \quad i \neq j, \qquad (2)$$

Table 2
K_d and K_c Constraint Statistics for Gupta's Log

Section	*9.1*	*9.2*	*9.3*	*9.4*	*9.5*	*ALL*
K_d	.500	.850	.650	.550	.500	.300
K_c	.625	.875	.875	.625	.500	.375

Note. The ALL graph is normalized relative to STUDY's theoretically maximal (canonical) graph. Graphs representing individual sections are normalized relative to the canonical ALL graph.

where x_{ij} is the entry in the i^{th} row j^{th} column of an adjacency matrix and $\frac{N^2-N}{2}$ is the number of possible pairs of nodes. Like density, cohesion ranges from a minimum of 0 to a maximum of 1. The more cohesive a graph, the less constraint a learner imposes on selections from the set of actions. By reflecting cohesion, we define a second measure of discriminated action, K_c. The value of K_c for Gupta's ALL graph is .50 normalized relative to STUDY's theoretically maximally cohesive graph. Gupta's rules disallowed two two-way relations (T→U and T→E) of 10 possible. Table 2 presents values of K_c, normalized relative to Gupta's set of rules, for each of the five sections of text.

By carefully defining a canonical graph relative to which empirically generated graphs are normalized, we suggest it is possible to characterize degrees to which a student exercises discriminated action in studying. Such discriminations depend on attention and conditional knowledge (Winne & Butler, 1994), that is, perceptions about the match between information in the instructional environment at a given time and the student's cognitive states. Based on the poststudying interviews with Gupta, we speculate that K_d may reflect the richness of a student's conditional knowledge about studying tactics to use while learning relative to the student's understanding of the task assigned.

Identifying Cognitive Tactics

Treatments in instructional research are designed to engage students in particular patterns of cognition triggered by particular conditions. For example, to promote the hypothetical cognitive activities of (a) focusing attention by using a rule to monitor information, (b) rehearsing important information, and (c) enhancing retrievability of that information by meaningfully elaborating it, students might be trained to (a) discriminate important information in a text, (b) underline it, and (c) make a note about it. These events comprise a cognitive tactic, a pattern among actions that a learner uses to achieve a relatively specific goal. Tactics can be represented as condition-action rules: if condition A is true (information is judged important to learn), then perform action X (reread that information as it is underlined); if condition B is true (rehearsal is

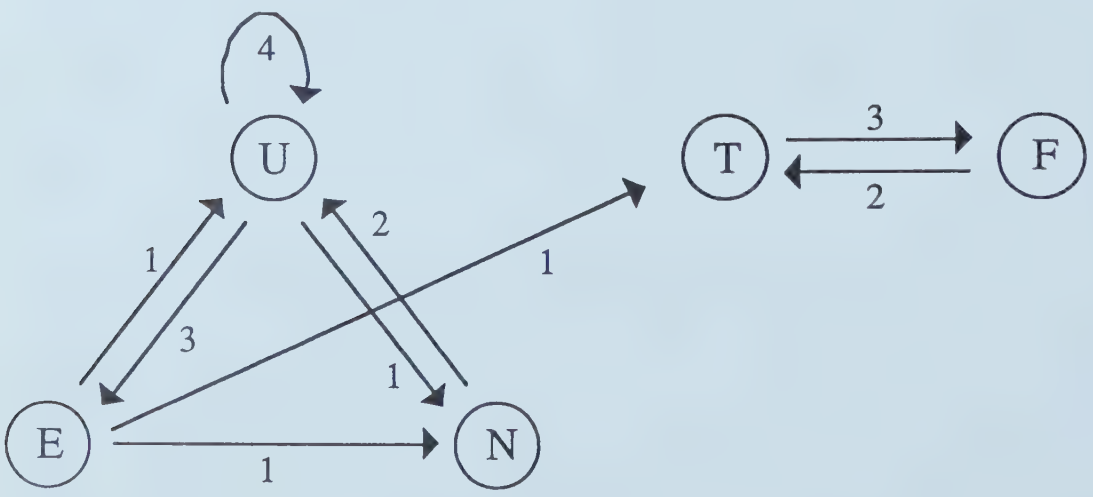


Figure 4. Graph of studying in Section 9.4 {U U U E N U N U E U U U E T F T F T F}.

complete and retrievability should be enhanced), then perform action Y (make information meaningful by paraphrasing it in a note). By equipping STUDY with tools for a student to trace each element in this cognitive tactic, a graph of those traces can represent the pattern of actions whenever the student applies that tactic. Graph theory provides tools that help to identify tactics within action sequences and that describe the degree to which tactics resemble one another or resemble a canonical pattern.

One defining feature of a cognitive tactic is that it is engaged only under particular conditions. This feature can be manifested in a graph of studying events by a node that is a gateway to a tactic graphed as a small, tightly linked pattern among other nodes. In the graph of Gupta's studying in Section 9.4 of the text (Figure 4), node E plays this role. A request for an elaboration led Gupta to engage in a reverberating cycle of asking for a test (T) question and requesting feedback about her answer (F).

In graph theory a *bridge* is a link in a connected graph that, if removed, creates two separate connected graphs. Bridges can represent links where the preceding act triggers a tactical response. In the graph for section 9.4, link E→T is a bridge. Removing that link creates two isolated graphs, one of which is a recognizably common satellite studying tactic, T↔F.

Another concept in graph theory, the *cut point*, also may reveal tactics. If a node in a connected graph is deleted, all links into and out from that node are deleted. A cut point is a node in a connected graph that, if deleted, causes the graph to become disconnected. In the graph of section 9.4 (Figure 4), there are two cut points, nodes E and T. Nodes at either end of a bridge often, but not always, are cut points.

Bridges and cut points may not always identify tactics; a "general utility" tactic might be embedded in a graph in a less detectable way. Such tactics might be identifiable by capitalizing on another description of graphs called *strong components*. To reach this concept, we must build on the definition of a *clique*.

A clique is a subgraph within a larger graph in which every node is linked to every other node.[3] That is, a clique is a subgraph in which K_c, one of our constraint statistics, has a value of 0. A strong component relaxes the definition of a clique to allow subgraphs with $K_c \geq 0$, but requires that each node connect directly to every other node or be connected through intermediate nodes within the strong component. Table 3 identifies strong components in the log of Gupta's studying. Studying in Sections 9.1 (see Figure 5) and 9.4 share two strong components, one a three-step tactic involving underlining, making notes, and requesting elaborations and another consisting of the testing-feedback tactic discussed above.

Describing Resemblance Among Actions

Operationally different actions a student performs in STUDY's tutoring environment may or may not trace different forms of cognitive processing. Consider the actions of underlining and note making. These seem distinct. However, each could trace the same cognitive processing, for example, a condition-action rule that discriminates important information followed by maintenance rehearsal enacted as either rereading while underlining or copying text verbatim to a note. If two observed actions can substitute cognitively

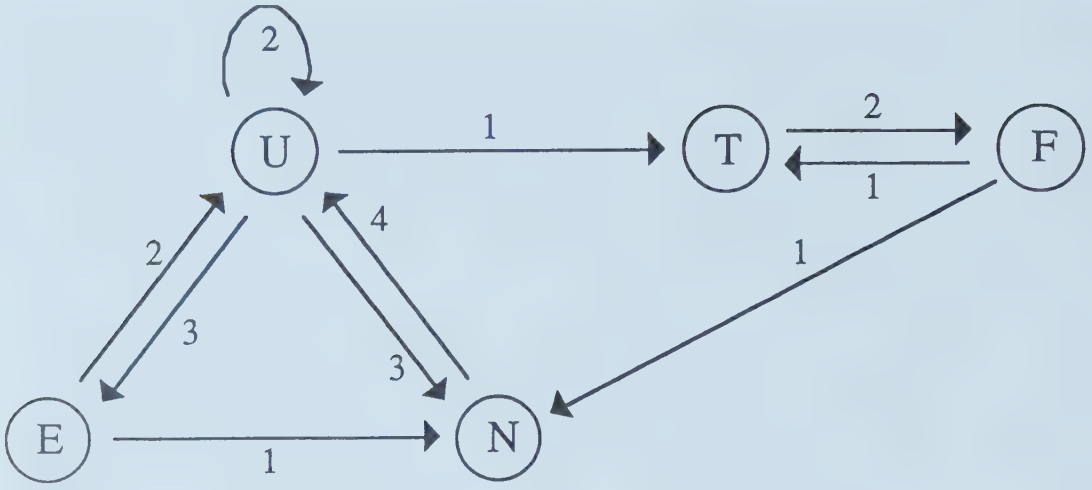


Figure 5. Graph of studying in Section 9.1 {U U N U E U E U N U N U U E N U T F T F N}.

for one another, and if the learner has no preference about alternate expressions for that cognitive processing, the nodes representing each action should have nearly identical relations to other nodes in a graph of the student's studying.

A graph theoretic statistic called structural equivalence can support inferences about whether operationally different traces reflect essentially the same cognitive processing. Structural equivalence compares the roles that two nodes play in a graph by measuring the pair's relationship to every other node in their graph. For each pair of nodes i and j, the distance between them is calculated relative to all other nodes, k, in the graph.

$$d_{ij} = \sqrt{\sum_{\substack{k=1 \\ k \neq i,j}}^{N} (x_{ik} - x_{jk})^2 + (x_{ki} - x_{kj})^2} \quad , \qquad (3)$$

where i, j, and k are nodes in a graph, and x_{ij} is the entry in the i^{th} row, j^{th} column of the graph's weighted adjacency matrix. The smaller the structural equivalence coefficient, the more two nodes resemble one another in terms of how those nodes relate to all other nodes. A limiting value of 0 indicates perfect substitutability.

Table 4 shows structural equivalence coefficients for each pair of nodes in Gupta's ALL graph computed using the weighted adjacency matrix. The statistics suggest that actions E (elaborations) and N (notes) function similarly with

Table 3
Strong Components in Gupta's Log

Section	*9.1*	*9.2*	*9.3*	*9.4*	*9.5*
Strong components	[E, U, N] [T, F]	[E, U]	[T, F, N]	[E, U, N] [T, F]	[E, U] [T, F, N]

respect to all the other nodes in the graph, as do T (generating a test question) and F (feedback). We interpret that E and N are both elaborations playing the same general role in Gupta's cognitive engagement, but they differ mainly in their source of information: E is an elaboration that STUDY would provide, whereas N is an elaboration the student generates. Regarding T and F, a similar interpretation is implied. By generating a question about material studied, Gupta creates internal feedback about qualities of her understanding. Feedback that STUDY provides differs in the origin of its information, but it serves essentially the same function as self-generating a question: both actions seek feedback to guide further studying.

A researcher may sometimes be interested in measuring the degree of disresemblance between actions. For instance, suppose a student is being trained to use a study strategy. Plotting resemblance statistics across training sessions describes intraindividual changes—development, functional fixedness, adaptation—in the student's cognitive processing as training progresses. Used in interindividual comparisons, resemblance statistics compared across students can characterize differences in cognitive processing routines.

Describing Resemblance Among Cognitive Tactics and Strategies

As well as comparing single actions that trace features of cognitive processing, research also needs to compare the degree of resemblance or congruence that tactics and strategies have to one another. This would allow comparing interindividual differences in tactics and strategies, as well as intraindividual changes in a student's tactics and strategies when studying different kinds of text or preparing for different kinds of examinations. The graph theoretic statistic multiplicity serves this purpose. For two (or more) graphs that have in common one set of possible nodes, multiplicity measures on a scale from 0 to 1 the number of links the graphs have in common normalized relative to the number of possible links,

$$M = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} x_{ij} \cdot y_{ij}}{N^2}, \quad i \neq j, \tag{4}$$

Table 4
Structural Equivalence of Nodes in the ALL Graph Representing Gupta's Studying

Using the weighted adjacency matrix

	E	*F*	*N*	*T*	*U*
E	.00	16.34	4.69	16.67	7.28
F		.00	13.89	2.24	17.55
N			.00	13.34	10.15
T				.00	17.52
U					.00

Table 5
Multiplicity Values for Sections in Gupta's Log

	9.1	9.2	9.3	9.4	9.5
Section 9.1		.15	.20	.40	.35
Section 9.2			.05	.15	.15
Section 9.3				.25	.30
Section 9.4					.30
Section 9.5					

where x_{ij} is the entry in the i^{th} row j^{th} column of the adjacency matrix for graph X, y_{ij} is the entry in the i^{th} row j^{th} column of the adjacency matrix for graph Y, and N is the number of nodes in each graph.

Again, the question arises, What canonical graph should be used to normalize the statistic? As discussed earlier, we replace N^2 in equation 4 with the number of followed-by links possibly observed in Gupta's data (N=20). Using this adaptation, multiplicity values presented in Table 5 compare the structure of all pairwise sets of graphs tracing Gupta's studying normalized relative to a completely unconstrained canonical ALL graph defined by Gupta's presumptions (i.e., 20 possible links).

The most similar graphs are those for Sections 9.1 and 9.4 (see Figures 4 and 5). Visually these graphs suggest that Gupta used nearly identical studying strategies and tactics. Only the links between the subgraphs [U, E, N] and [T, F] differ across these two sections of the text. Despite differences in the frequency of individual studying acts, there is strong resemblance of Gupta's studying across these sections.

We also want to directly compare pairs of graphs having a common set of possible nodes to one another rather than normalizing them relative to a canonical graph. The question arising in this situation is how to define the normalizing term in equation 4. Our proposal is to sum over the pair of graphs the number of distinct links observed.

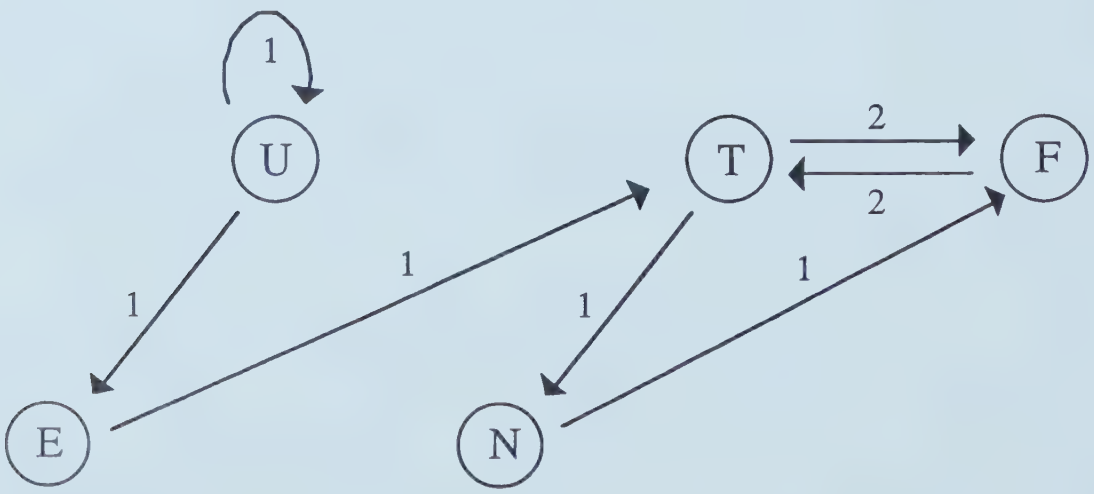


Figure 6. Graph of studying in Section 9.3 {U U E T N F T F T F}.

Table 6
Normalized Multiplicity Values for Gupta's Log

	9.1	9.2	9.3	9.4	9.5
Section 9.1	1.00	$\frac{3}{10}$=.30	$\frac{4}{13}$=.37	$\frac{8}{11}$=.73	$\frac{7}{13}$=.54
Section 9.2		1.00	$\frac{1}{9}$=.11	$\frac{3}{9}$=.33	$\frac{3}{10}$=.30
Section 9.3			1.00	$\frac{5}{11}$=.45	$\frac{6}{11}$=.55
Section 9.4				1.00	$\frac{6}{13}$=.46
Section 9.5					1.00

By the first measure of graph multiplicity in Table 5, Sections 9.3 (see Figure 6) and 9.5 were as comparable as Sections 9.4 and 9.5, M=.30 in both cases. By our alternative measure, Sections 9.3 and 9.5 are more similar to each other, M=.55 versus M=.46 comparing Sections 9.4 and 9.5. This variation is explained by relationships among T, F and N: these actions interact partially in Section 9.3 but not at all in Section 9.4. We intuit that our adapted measure of graph multiplicity is a better gauge of the similarity of graphs and, correspondingly, how Gupta studied in different sections.

Conclusion

We propose that, fundamentally, individual differences exist as cognitive acts that learners perform over the course of engaging with a task. Many, perhaps most, cognitive acts are difficult for a learner to inspect, report, or otherwise reveal because they are: complex, automatic cognitive procedures that can be decomposed only with difficulty and at cost to smooth and effective execution (Anderson, 1983; McKoon & Ratcliff, 1992); tacit (Schommer, 1990; Winne & Butler, 1994); or simply inexpressible with terms in the learner's lexicon. Facing these problems, traditional approaches to studying individual differences such as questionnaires about learning styles and measures of global achievement have aggregated up. Such data obscure individual differences at their source.

A decade ago, Winne (1982) argued that data representing individual differences should be gathered at their source level and examined at that level

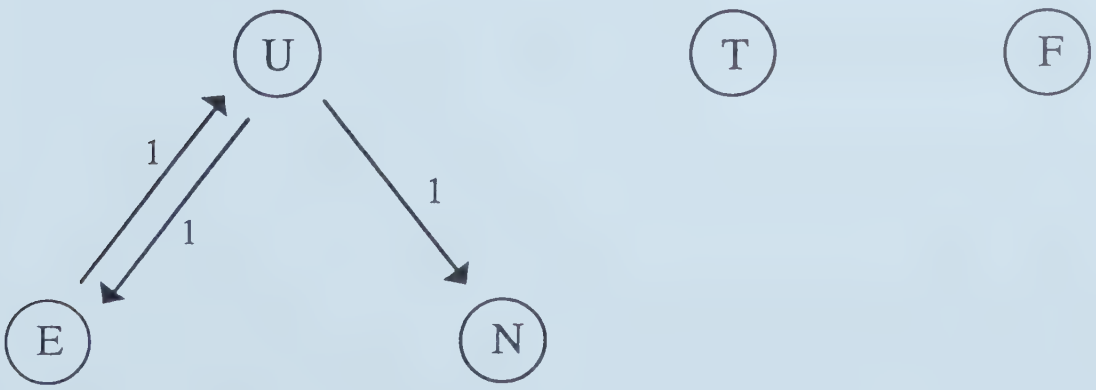


Figure 7. Section 9.2 {U E U N}.

before aggregating up. He argued that analyzing fine-grained traces of cognition would sharpen pictures of how learners express intra- and interindividual differences, and provide more penetrating grounds for examining how individual differences affect the course of learning and its ultimate achievements. The adaptive learning system STUDY was designed in part as a tool for gathering trace data. The question remaining, however, was: How can trace data be analyzed to reveal cognitive acts and patterns that constitute individual differences?

Graph theoretic concepts are one source of techniques for addressing this question. In a task like studying, descriptors and statistics introduced in this article are used to (a) identify units of cognition and patterns of cognitive engagement, and (b) measure the degree to which traces of single cognitive units as well as cognitive tactics resemble one another. Coupled with traces of cognition that are carefully designed to reflect theoretically important cognitive events, these measures can characterize features that give rise to individual differences, namely, conditional knowledge, attention, discriminated action, and goals.

It is important to recognize limitations in this approach to investigating individual differences in tasks such as studying that entail complex cognition. A sequence of traces reflects not only knowledge and competence a learner brings to task, but also the interaction between these and the information studied. For example, the number of study actions and the pattern of studying in Section 9.2 (see Figure 7) suggest that Gupta was quite inactive relative to other sections. Does this indicate lack of comprehension, perhaps occasioned by low prior knowledge, or low levels of strategy use? Section 9.2 was an extended example of applying statistical techniques covered in previous chapters of the text. Although graph theoretic concepts and statistics (see Tables 2 and 3) identified a difference in studying, other information about what was being studied must be considered to grasp how the difference makes sense. This reflects an important principle, that interpreting a statistic is inherently bound up with what raw data reflect. We continue to explore means by which to accommodate the situational qualities of trace data. Although this will complicate matters, we expect that fine-grained studies of such person-environment interactions will also spur newer and more valid theories of individual differences called for by Corno and Snow (1986).

Acknowledgment

This research was supported by a grant to the first author from the Social Sciences and Humanities Research Council of Canada (#410-92-0239).

Notes

1. These data cannot be regarded as providing a sufficient basis for theorizing about actual study strategies.
2. Presumptions about useful or permissible studying activities correspond to decisions a designer would have option to operationalize as rules governing a STUDY tutorial. For example, STUDY can be designed to provide feedback in relations other than a one-to-one correspondence to a specific test question. A designer might supply feedback to a student whenever a particular studying action is carried out, such as alerting the student that text just underlined is a minor detail rather than contributing to the text's main theme or message.
3. The requirement that a node be related to other nodes insures that a node with a self-loop cannot create a clique of size one.

References

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Butler, D., Winne P.H., & McGinn, M. (n.d.). *A model of feedback in the self-regulation of learners*. Unpublished manuscript.

Corno, L., & Snow, R.E. (1986). Adapting teaching to individual differences among learners. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 605-629). New York: Macmillan.

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Genest, M., & Turk, D.C. (1981). Think-aloud approaches to cognitive assessment. In T.V. Merluzzi, C.R. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 233-269). New York: Guilford Press.

Glover, J.A. (1989). The testing phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*, 392-399.

Goldman, S.R., & Saul, E.U. (1990). Flexibility in text processing: A strategy competition model. *Learning and Individual Differences, 2*, 181-219.

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review, 3*(1), 5-39.

Guzdial, M. (1993). *Deriving software usage patterns from log files*. Unpublished manuscript.

Haller, E.P., Child, D.A., & Walberg, H.J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher, 17*(9), 5-8.

Howell, D.C. (1987). *Statistical methods for psychology* (2nd ed.). Boston: PWS-Kent.

Johnson, L.L. (1988). Effects of underlining textbook sentences on passage and sentence retention. *Reading Research and Instruction, 28*(1), 18-32.

Kierwa, K.A., DuBois, N.F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*, 240-245.

Luchins, A.S., & Luchins, E.H. (1959). *Rigidity of behavior: A variational approach to the effects of einstellung*. Eugene, OR: University of Oregon Books.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440-466.

Peterson, S.E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction, 31*(2), 49-56.

Schommer, M. (1990). Effects of belief about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*, 489-504.

Walczyk, J.J., & Hall, V.C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology, 81*, 435-437.

Winne, P.H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science, 11*, 13-28.

Winne, P.H. (1992). State-of-the-art instructional computing systems that afford instruction and bootstrap research. In M. Jones & P.H. Winne (Eds.), *Foundations and frontiers of adaptive learning environments* (pp. 349-380). Berlin: Springer-Verlag.

Winne, P.H. (1993a, May). *Frontiers of research on academic learning*. Paper presented to the Canadian Psychological Association, Montreal.

Winne, P.H. (1993b). A landscape of issues in evaluating adaptive learning systems. *Journal of Artificial Intelligence in Education, 4*, 295-318.

Winne, P.H., & Butler, D.L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.). Oxford, UK: Pergamon.

Winne, P.H., Jones, M., & Field, D. (1992, June). *STUDY: A tool for authoring adaptive learning environments and for advancing instructional research*. Paper presented to the 2nd International Conference on Intelligent Tutoring Systems, Montreal.

Winne, P.H., Jones, M., Field, D., & Nesbit, J.C. (n.d.). *Instructional research with STUDY*. Unpublished manuscript.

Zimmerman, B.J., & Schunk D.H. (Eds.). (1989) Self-regulated learning and academic achievement. Springer-Verlag: Berlin

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 195-211*

W. Todd Rogers
University of Alberta

and

David J. Bateson
University of British Columbia

Verification of a Model of Test-taking Behavior of High School Seniors

Thirty-six test-wise and 41 test-naive high school seniors were individually tested using a 14-item test designed to assess their abilities to apply selected test-wiseness principles (Millman, Bishop, & Ebel, 1965) according to a proposed model of test-taking behavior. Students' approaches to their answers were recorded and analyzed. Overall, the results (a) suggest that before students can profitably apply test-wise skills, they must first possess knowledge about the content of the stem and/or options to eliminate an option or options or to take advantage of item cues, and (b) support the proposed model of test-taking behavior.

Trente-six élèves du secondaire ayant développés l'art d'écrire des examens et 41 élèves du même niveau n'ayant pas maîtrisé cet art subirent des épreuves individuels au moyen d'un test à 14 items. Ce test a été conçu pour évaluer leurs habiletés à appliquer certains principes de l'art d'écrire un examen (Millman, Bishop, & Ebel, 1965) d'après un modèle proposé pour déterminer le comportement des élèves durant un examen. La façon que les élèves abordaient les réponses fut notée et analysée. En somme, les résultats suggèrent (a) qu'avant de pouvoir utiliser profitablement les habiletés et le savoir faire pour répondre aux questions d'un test, les élèves doivent d'abord posséder une certaine connaissance au sujet du contenu de la souche de la question et/ou des options pour éliminer une option ou des options ou pour utiliser des indices inhérentes à la question et (b) que les résultats appuient le modèle proposé des comportements en testing.

> STANDARD 3.11—When test-taking strategies that are unrelated to the constructs or content measured have been found to influence test performance significantly, these strategies should be explained to test takers before the test is administered either in an information booklet or, if the explanations can be made briefly, along with the test directions. The use of such strategies by all test takers should be encouraged if their effect facilitates performance and discouraged if their effect interferes with performance. (*Primary*). (American Psychological Association, 1985, pp. 27-28)

Test-wiseness is a cognitive ability or set of skills that a test taker can use to improve a test score no matter what the content area of a test (Benson, 1988;

Todd Rogers is a professor in the Department of Educational Psychology and Director of the Centre for Research in Applied Measurement and Evaluation. He specializes in measurement and evaluation.
David Bateson is an associate professor in the Department of Mathematics and Science Education and a research associate with both the Educational Measurement Research Group and the Centre for Applied Studies in Evaluation. He specializes in science education, measurement, and evaluation.

Sarnacki, 1979). If a test taker possesses test-wiseness, and if the examination contains susceptible items, then the combination of these two factors can result in an improved score; in contrast, a student low in test-wiseness will tend to be penalized every time he or she takes a test that includes test-wise components.

A number of authors have offered definitions of test-wiseness (Diamond & Evans, 1972; Ebel & Damrin, 1960; Gibb, 1964; Stanley, 1971; Thorndike, 1951; Vernon, 1962), but for the purposes of this study the definition proposed by Millman, Bishop, and Ebel (1965) was adopted: "a subject's capacity to utilize characteristics and formats of the test and/or test taking situation to receive a high score" (p. 707). Basically, then, test-wiseness encompasses both the method of measurement (flawed test items that provide test-wise cues) and characteristics of the test taker (a cognitive ability or set of abilities that an examinee might employ in any testing situation regardless of the content measured).

To further explicate the construct of test-wiseness, Millman et al. (1965) included with their definition a taxonomy of test-wiseness principles that has served as the general framework for further studies of test-wiseness. Briefly, this taxonomy is organized into two categories. Part I contains elements applicable in most testing situations and that are independent of the test maker or test purpose. If employed, these strategies will help examinees avoid losing points for reasons other than lack of knowledge of the content tested. The principles listed in Part II of the taxonomy may prove beneficial when the test taker has knowledge of particular test making behaviors or knowledge of particular testing practices gained from past experiences with tests similar in purpose and format. For more detail, the reader is referred to Millman et al., (1965) and Sarnacki (1979).

Model of Test-wise Test Taking Behavior

Displayed in Figure 1 is a proposed model of the test taking behavior of skilled test takers. Based on the work of Brown (1980; 1987), Flavell (1979), and Schuell (1986), and an earlier model proposed by Smith (1980), the model reflects various routes a skilled examinee may take to determine what option to select on a multiple-choice item, and an "executive" governor for selecting which routes are used. Briefly, the model suggests that the cognitions include:

1. a cognitive monitor that controls which abilities and skills are going to be engaged to answer the item under consideration;
2. knowledge, abilities, and skills relevant to the content or trait being measured;
3. knowledge of test-wiseness principles; and
4. the response (selection and record of choice).

The proposed process follows a defined path. First, the test taker reads the stem of a multiple-choice item and then attempts to recognize, using knowledge about the perceived content being tested, what he or she believes to be the correct answer from among the options listed. If the answer is not found, an unskilled test taker will either simply guess from among the options presented or omit the question entirely. In contrast, a skilled test taker, by way of his or her cognitive monitor for testing and partial knowledge about the content being measured (including that contained in the item's options), will next apply the set of test-wiseness principles he or she possesses, working cyclically through the elements of the set for a test-wiseness element-item cue match.

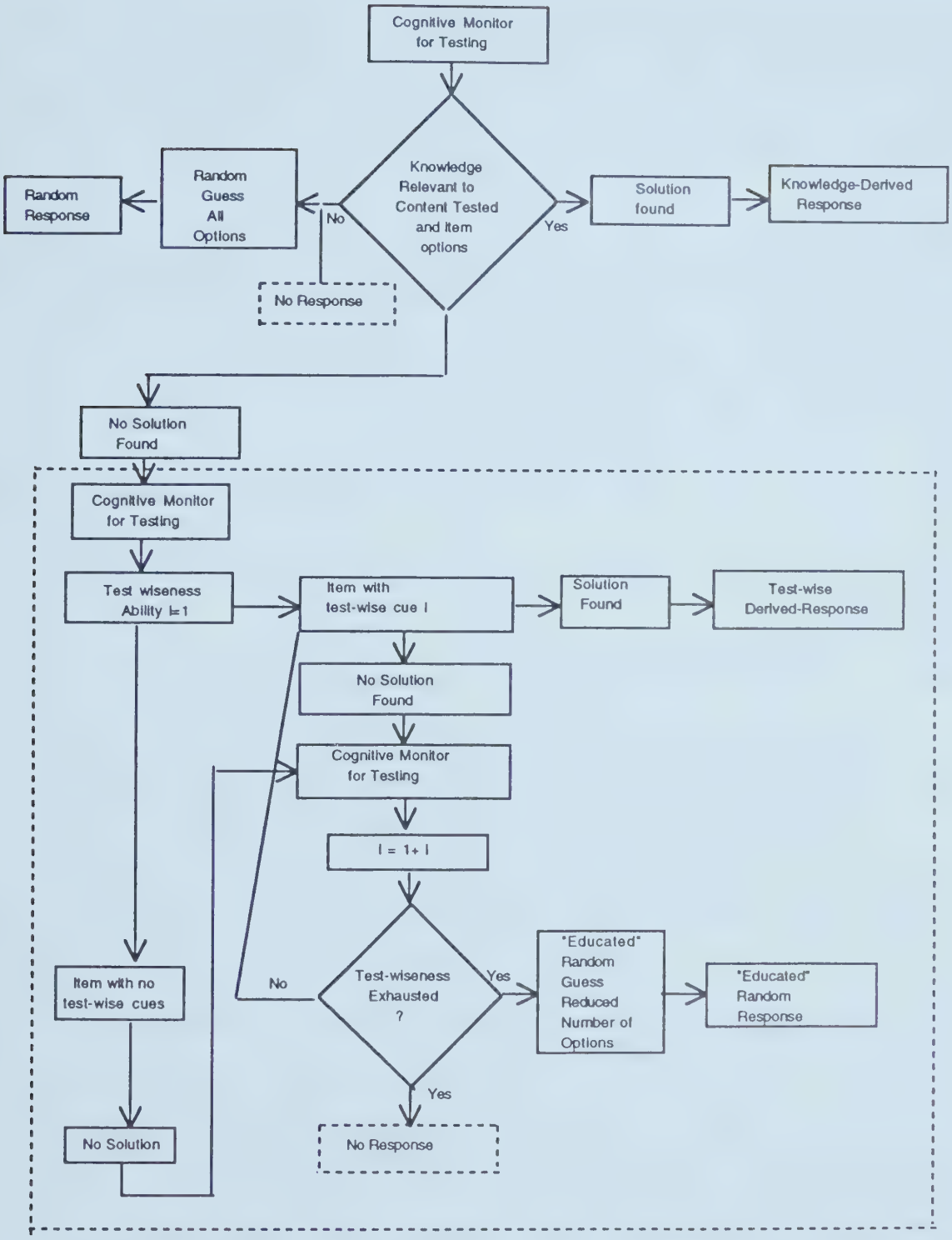


Figure 1. Proposed model of test-taking behavior of skilled test-takers.

When a match is made, the cycle is terminated and a test-wise (as opposed to a pure knowledge) response is recorded. In the case of no match, either because there are no item test-wise cues or because the test-taker has exhausted all of his or her test-wiseness strategies, the skilled test taker will probably make an "educated" random response. Clearly then, according to this model, other characteristics of the skilled test taker in addition to knowledge of the content being tested enter into a high test-wise student's final test score.

The study described and reported in this article was designed to assess the validity of this test-taking model. Completed as part of a larger study of the

impact of test-wiseness on the performance of high school students in British Columbia on school leaving examinations (Rogers & Bateson, 1991), the intent was to gain a more complete understanding of test-wiseness and the way and extent to which high school seniors actually employ test-wiseness when responding to multiple-choice items.

Method

Test of Test-wiseness

The Test of Test-wiseness (TTW) developed and used in this study was divided into two sections. Section 1 contained 24 four-option, multiple-choice items evenly distributed across four of the test-wiseness elements listed in the taxonomy outlined by Millman et al. (1965). The four elements included:

1. Three deductive reasoning strategies:

 ID1—eliminate options known to be incorrect;

 ID2—choose neither or both of two options which imply the correctness of each other;

 ID3—choose neither or one of two-options, one of which, if correct, would imply the incorrectness of the other; and

2. One cue-using strategy:

 IIB4—recognize and use similarities between the stem and the options.

The selection of these four elements was based on the most frequently occurring item faults identified in the provincial examinations administered during the previous year in British Columbia. The 24 items used were selected from test-wiseness tests developed earlier by Gibb (1964), Millman (1966), and Slakter, Koehler, and Hampton (1970). To provide face validity, the content of the items, unfamiliar to the students, was distributed approximately evenly across the content areas of four grade 12 courses with the largest student enrollments—English, algebra, history, and biology (see Table 1). However, the actual content tested by each item was deliberately unfamiliar to the students so as to rule out the possibility that correct answers would be attributable to student knowledge alone.

The intent of the first section was to measure the degree to which students were able to use deductive reasoning to answer questions about content unfamiliar to them, but that were flawed in some way. Group administered, the instructions to the students were:

> This is a test of test-wiseness which measures some of the abilities needed to do well on tests. Many of the questions are about things you may not have studied. However, there are test-taking strategies which can be used to figure out what to do when faced with such questions.
>
> *For example:*
> *The greatest advantage of using slent in the manufacture of steel is that slent makes steel*
>
> a. transparent.
> b. stainless.
> c. heavy.
> d. rubbery.

Table 1
Description of Test of Test-wiseness

Test-wiseness Principle	*Content Area*				
	English	*Math.*	*History*	*Biology*	*Total*
Section 1					
Absurd options(1D1)	2	1	2 (2)	1	6(2)
Similar options(1D2)	2	1	1 (1)	2 (1)	6(2)
Opposite options(1D3)	2	2	1	1 (2)	6(2)
Stem-option (IIB4)	1	1 (1)	2 (1)	2	6(2)
Total	7	5 (1)	6 (4)	6 (3)	24(8)
Section 2					
True-False	1	- (1)	-	1 (1)	2(2)
Multiple-Choice					
Guessing	1	1 (1)	1 (1)	1	4(2)
Non-guessing	1	1	1 (1)	1 (1)	4(2)
Total	3	2 (2)	2 (2)	3 (2)	10(6)

Note: Numbers in parentheses refer to the number of items on the interview form.

> Using test-wiseness strategies, options 'a' and 'd' can be eliminated because they are clearly not correct (steel is not transparent, nor is it rubbery). Therefore, either 'b' or 'c' is the correct answer. Now we stand a better chance of guessing the correct answer for we have narrowed the number of possible options down to two from four.
>
> *Please be sure to follow the specific instructions for each of the following sections.*
>
> *Section 1: Suggested Time: 20 minutes*
>
> *INSTRUCTIONS: For each question, select the BEST answer and record your choice on the answer sheet provided. Each question is worth one mark. There will be no correction for guessing.*

Section 2 involved the use of guessing whenever the chance of profiting was positive. Following Millman (1966), the penalty system for this section was such that test-wise behavior should lead examinees to guess on true-false questions and those multiple-choice questions for which more than one option could be eliminated:

> INSTRUCTIONS: For each question in this part of the test, select the BEST answer and record your choice on the answer sheet provided.
>
> Each question answered correctly is worth 5 points. Zero points will be given for each omitted question. Two points will be *deducted* for each question answered incorrectly.

Of the eight multiple-choice questions asked, four were such that guessing was profitable; the options on the remaining four questions were such that it was

not possible to clearly eliminate an option (see Table 1). The following two examples illustrate these two types of questions:

Example 1. Guessing is profitable

30. The human skin
 a. possesses a Malpighian layer.
 b. is always brown.
 c. cannot regenerate.
 d. is uncolored.
 e. is composed of cuboidal epithelium.

Example 2: Guessing is not profitable

32. What is the value of 1n sin 8º15′ to five decimal places?
 a. .14321
 b. .14332
 c. .14349
 d. .14352
 e. .14361.

Scoring. The system used to score the TTW was designed to reflect the use of the test-wiseness elements assessed. Modelled after Millman (1966), the items were scored as follows:

Section 1. One point for

ID1: correct answer;
ID2: either of the two nonsimilar options;
ID3: either of the two opposite options; and
IIB4: the cued option.

Thus the range of possible scores was zero to 24; a high score reflected frequent use of test-wiseness reasoning.

Section 2. Two points were awarded for every true-false and "guessing" multiple-choice item answered, regardless of whether the answer was correct. Three points were awarded for each of the "nonguessing" multiple-choice items not answered. Thus the maximum score was 24; a high score reflected appropriate use of a guessing strategy.

Interview form. A second, shorter form of the TTW was developed and administered on an individual basis to identify the actual processes students follow when answering questions with unfamiliar content. This form consisted of eight four-option, multiple-choice items evenly distributed across the four test-wiseness elements included in the longer, group administered form and two true-false and four five-option multiple-choice items to measure guessing strategy (see Table 1). Questions were obtained from Millman (1966) and the provincial biology examination administered the previous year.

Subjects

The subjects were students who wrote the provincial examination set for English 12 in June 1989. All grade 12 students must write either English 12 or Communications 12 in order to graduate. English 12, intended for students who are planning or wish to leave open the opportunity to pursue some form of tertiary education, is more popular, enrolling between five and six times as many students as Communications 12. Further, students who take English 12

also elect additional courses with provincial examinations so as to meet postsecondary entrance requirements. Thus, by selecting students in English 12, samples of adequate size would be realized for each of the five remaining provincial tests considered in the larger study.

The students attended 10 public schools selected to provide provincial representation. Three of the schools—two large and one medium sized—were located in the southwest region of British Columbia. The others were distributed throughout the province: one large, one medium, and one rural school in the south central region; one medium and one rural school in the north central region; and two rural schools on the north coast. The students attending these schools were from diverse socioeconomic backgrounds, reflected the gender distribution and major ethnic groups in the province, and represented the full range of achievement. Summarized in Table 2 are the gender and ethnic distributions for the English 12 student sample (n=736) and the corresponding population distributions. Comparison of the sample values with the corresponding population parameters confirms the representativeness of the student sample (see Table 2).

Table 2
Gender, Age, and Ethnicity

Variable	*English 12 Sample (N=736)* *f*	*p*	*Population Percentage*
Gender			
Male	343	46.4	47.9[a]
Female	393	53.2	52.1
Age (Year of Birth)			
1967	1	0.1	n.a
1968	4	.5	n.a
1969	5	.7	n.a
1970	87	11.8	n.a
1971	618	84.0	n.a
1972	14	1.9	n.a
Ethnicity			
English	430	58.2	61.8[b]
French	21	2.8	2.4
Indigenous	11	1.5	2.1
East Indian	30	4.1	n.a.
Chinese	36	4.9	4.0
German	44	6.0	5.2
Italian	24	3.2	1.6
Japanese	8	1.1	n.a.
Other	132	17.9	n.a.

Note: Percentages do not total 100% due to omitted responses. n.a. = not available.
[a]Percentage of total Grade 12 provincial population.
[b]Percentage of total provincial population in 1986 (Sullivan, 1988, p.27).

Data Collection

The student data considered in this study were collected in two stages. At stage 1, an initial sample of 936 students in the 10 schools completed the TTW during an English 12 class at the beginning of May. The TTW was administered by the authors assisted by the teachers when more than one English 12 class was scheduled at the same time. The testing time required was 25 minutes.

Approximately two weeks later a subsample of the highest scoring (test-wise; n=36) and lowest scoring (test-naive; n=41) students on the nonguessing subtest of the TTW were individually tested to ascertain the response strategies the students used when answering each item in the interview form of the TTW. In order to reduce costs, this subsample was restricted to students enrolled in the six schools located in the southwest and south central regions of the province. The students were asked by the interviewer, a member of the research team, to "think aloud," to "describe and explain what you are doing," or "how did you get your answer?" for each of the 14 questions. Each interview lasted 15 to 20 minutes. Generally, the test-wise students required less time than the test-naive students.

Results

Description of Interview Samples

As previously described, the subjects selected for this part of the study formed two subsamples: 36 test-wise students whose scores on Section 1 of the TTW were greater than 17, and 41 test-naive students whose scores on Section 1 of the TTW were less than 11. The cut-off scores are approximately one standard deviation above and below the mean on Section 1 of the TTW for the initial sample of 936 students ($\overline{X}$ = 13.9; s=2.80). The percentages of males and females in both groups were approximately equal: 38.9% of the test-wise sample and 41.5% of the test-naive sample were male, and 61.1% of the test-wise sample and 58.5% of the test-naive samples were female. Nearly equal percentages of students in both groups, 52.8% and 53.7%, reported that they were of British/European ancestry. The next most frequently reported ethnic background was "Other," 36.1% and 19.5%. The difference between these two percentages is mainly accounted for by smaller percentages of Aboriginal (0.0% vs 4.9%) and East Indian (5.6% vs 12.2%) students in the test-wise sample than in the test-naive sample.

Solution Strategies Used

Turning now to the strategies used by the students in these two samples, the percentages of students in each sample who correctly responded to each item included in Section 1 of the Interview Form are reported in Table 3.

The results are organized in terms of the strategies that, if used properly, would lead to an increased probability of a correct response. Table 3 reveals that the differences between item percentages in favor of the test-wise sample are significant on two items (4 : z = 2.65; 8 : z = 2.35; $p < .05$). Despite the lack of statistical significance on other items in Section 1 with large differences (items 1, 2, 5, 6, 7), there are discernible differences between the solution strategies the test-wise students reported they used and the solution strategies the test-naive students reported they used. These are described next in terms of test-wise principles most frequently used by the students when answering each item.

Table 3
Performance of Interview Samples on Interview Form, TTW
Section 1

Test-wiseness element	Item #	Sample: Test-wise (n = 36) n	%	Sample: Test-naive (n = 41) n	%
ID1	1	19	52.8	14	34.1
	8	15	41.6	6	14.6*
ID2	4	27	75.0	20	48.8*
ID3	2	16	44.4	14	34.1
	5	25	69.4	22	53.6
IIB4	3	19	52.8	21	51.2
	6	28	77.8	25	61.0
ID1, ID2	7	27	75.0	29	70.7

*$p < .05$

ID1: Option known to be incorrect

1. The compromise between Democrats and Republicans after the post-civil war election of 1876 resulted in
 A. federal aid to Southern railroads.
 B. the entrance of Texas into the Union as a slave state.
 C. a treaty with Joseph Stalin.
 D. the Fugitive Slave Law. (Millman, 1966)

8. The emperor of the ancient Hsin Dynasty who resigned to undertake radical reform was
 A. Saigon.
 B. Wang Mang.
 C. Mao Tse Tung.
 D. Alexander I. (Millman, 1966)

Items 1 and 8 contain incorrect options that, if known by a student to be incorrect, should be eliminated or ignored. Ten (27.8%) students in the test-wise sample identified as wrong and then eliminated the three incorrect options in item 1. An additional four students first said they knew option C was wrong and then pointed out that, because B and D were similar, they too were wrong; a fifth test-wise student first indicated B and D were similar and then correctly guessed A. In contrast, five (12.2%) students in the test-naive sample correctly identified and eliminated the three incorrect options; no student in this sample reported B and D to be similar. Thirteen (36.1%) students in the test-wise sample and nine (22.0%) students in the test-naive sample eliminated one or two options (including, in some cases, A) as incorrect and then guessed among the remaining three or two options. Two (5.6%) students in the test-wise sample reported that they knew the answer (both incorrect), and seven (19.4%) indicated that they simply guessed (two correctly, which agrees with what would be expected by simple chance). The incidence of knowledge and guess-

ing in the test-naive sample was greater: 12 (29.3%) students said they knew the answer (only three were correct), and 15 (36.6%) reported they guessed (three correctly which again agrees with what would be expected by chance alone).

The pattern of solution strategies used by the students to answer item 8 is somewhat similar to that observed for item 1. Although 14 (38.9%) students in the test-wise sample correctly identified and eliminated the three incorrect options in item 8, only three (7.3%) students in the test-naive sample did likewise. Twelve (33.3%) test-wise students and 16 (39.0%) test-naive students eliminated one or two options (including in some cases the correct option) and then guessed among the remaining three or two options. Three additional students, one test-wise and two test-naive, used strategy ID1 to eliminate three options, one of which was the correct option. Two test-wise students first used ID1 to eliminate option D and then selected A by guessing (one student) or citing knowledge ("I know this from a class I took"). Four test-naive students followed the same strategy to eliminate A and/or B and then selected C (3) or D (1) citing knowledge. Two students in the test-wise sample and three in the test-naive sample used the stem-option similarity to select C. Three students in each sample selected their answer (all of which were incorrect) based on knowledge they claimed they had. Last, whereas only one test-wise student said he guessed, eight (19.5%) students in the test-naive sample said they guessed (two correctly).

ID2: Similar Options

4. The sigma effect of the Fahraeus-Lindgvist phenomenon is related to
 A. the flow of liquid through the kidney.
 B. the muscular contractions of the kidney.
 C. the kidney's transmission of fluids.
 D. the diameter of the red cells flowing through the kidney's blood vessels. (Millman, 1966)

As was the case for the two previous items, a greater percentage of students in the test-wise sample than the percentage of students in the test-naive sample correctly took advantage of the existence of similar options in item 4. Twenty-one (58.3%) test-wise students identified A and C as similar and selected their answers from B and D: 16 guessed, two used ID1 to eliminate D, another used ID1 to eliminate B, and two selected D based on knowledge "learned in class." In contrast, five (12.2%) test-naive students identified A and C as similar: four reported they guessed between B and D and one selected D based on previously learned knowledge. The majority of test-naive students, 25 (61.9%), reported they guessed (seven correctly); only nine (25.0%) test-wise students reported they guessed (two correctly). And whereas eight (19.5%) test-naive students said they knew the answer, only one test-wise student said she knew the answer.

ID3: Opposite Options

2. If ribulose biphosphate were removed from a chloroplast, which of the following statements would BEST describe the immediate result?
 A. CO_2 could not enter the Calvin cycle.
 B. ATP could not be produced in the thylakoid.
 C. O_2 could not enter the Calvin Cycle.
 D. Light energy could not be trapped in the grana. (Ministry of Education, 1986)
5. Which of the following would cause an oat seedling to bend to the right?
 A. Auxins placed on the left side of the shoot tip.
 B. Giberellins placed on the left side of the shoot tip.
 C. Cytokinins placed on the right side of the shoot tip.
 D. Auxins placed on the right side of the shoot tip. (Ministry of Education, 1986)

Unlike the previous three items, items 2 and 5 were selected from the previous year's provincial examination in biology. Consequently, although both items contained opposite options, several students, particularly in the test-wise sample, cited the use of previously learned material in arriving at their answer. In contrast, the incidence of guessing was greater in the test-naive sample. Further, a greater number of students used more multiple strategies on items 2 and 5 than they did on the three previous items. On item 2, for example, 20 (55.6%) test-wise students identified A and C as opposites, although only six (14.6%) test-naive students recognized that A and C were opposites. Among the 20 test-wise students, nine first indicated A and C were both incorrect and then selected B or D using knowledge (5) or guessing (4); three first eliminated option D as incorrect, recognized A and C were opposites and then selected A or C using knowledge (2) or guessing (1); and eight recognized A and C as opposites and selected between these options using knowledge (3) or guessing (5). Two of the six test-naive students who said A and C were opposites indicated both were wrong and, citing knowledge learned in class, selected B; one first eliminated D and then guessed between A and C; and three guessed between A and C. Although four (11.6%) test-wise and six (14.6%) test-naive students indicated they used knowledge alone to get their answers, only five (13.9%) test-wise students compared to 22 (53.9%) test-naive students simply guessed their answers.

As might be expected, the pattern of solution strategies used by the students to answer question 5 is similar to that for item 2. Again the majority of test-wise students, 26 (72.2%), recognized that A and D were opposites whereas a minority of test-naive students, 13 (31.7%) did the same. Six of the 26 test-wise students indicated that A and D were wrong: one then eliminated B as incorrect, another used the similarity between the stem and option C (the word "right"), three used knowledge learned in class to pick between B and C, and one, after omitting A and D, guessed between B and C. The remaining 20, all of whom selected either A or D, did so in the following ways: one first eliminated B, recognized A and C were opposites, and then guessed; two used the presence of "right" in the stem, one to select D, the other to select A; and 17 chose between A and C on the basis of knowledge (6) or guessing (11). Of the

13 students in the test-naive sample who recognized A and D were opposites, six indicated these options were incorrect and then selected option C (two used knowledge; one eliminated A as incorrect, one used stem-option similarity, and two guessed); one eliminated C as incorrect, recognized A and D as opposites, and then eliminated D as incorrect; and seven first recognized A and D as opposites and then selected A or D using knowledge (2), stem-option (1), or guessing (3). Again, whereas four (11.1%) test-wise students and six (14.6%) test-naive students indicated they used knowledge alone, only four (11.1%) test-wise students compared to 20 (51.2%) test-naive students simply guessed.

II.B.4: Stem-Option Similarity

3. In 1991 the Tories in control of the House of Commons showed their political influence and democratic spirit when they
 A. enacted the Asquith price-control bill.
 B. vetoed a bill to extend the suffrage privilege.
 C. forced the House of Lords to accept a reduction in its power.
 D. were defeated in an attempt to enact the National Insurance Act. (Millman, 1966)

6. Why is Cavalieri's Principle important in Solid Geometry?
 A. It shows that the surface area of a square of side s is s^2.
 B. It provides contradictions to the principles of Euclid and Gauss.
 C. It is used to prove that two polygons are congruent.
 D. It provides the basis for finding the volume formulae for many solids. (Millman, 1966)

As with the three previous test-wise elements, students in the test-wise sample made greater use of the stem-option cue built into items 3 and 6 than did students in the test-naive sample. Sixteen (44.4%) test-wise students identified a similarity between the stem and one of the options in item 3; 13 (31.7%) test-naive students did the same. Eleven of the 16 test-wise students correctly used the cue to select option C: seven directly, three following elimination of option D as incorrect, and one supported with knowledge learned in class. The remaining five students selected option B, one directly, one following elimination of D, and three supported by knowledge. Of the 13 test-naive students who used the stem-option cue, 10 used it correctly: nine directly and one following elimination of option D. The remaining three students all selected B, one directly and two with knowledge. Nine (25.0%) additional test-wise students used ID1 in determining their answer in much the same way as that observed for items 1 and 8: four students identified A, B, and D as wrong while two others identified B, C, and D as wrong; two identified A and D as wrong and then guessed between B and C; and one eliminated D and then guessed. Four (9.8%) test-naive students followed the same procedures: one eliminated A, B, and D to obtain the correct answer; one eliminated B and D and then guessed; and two eliminated D and then guessed. Eight (22.2%) test-wise students reported they simply guessed their answers (three correctly) in contrast to 18 (43.9%) test-naive students who reported they guessed their answers (eight correctly).

For item 6, 21 (58.3%) students in the test-wise sample used the stem-option similarity cue to obtain their answers, whereas 15 (36.6%) test-wise students

did likewise. Thirteen test-wise students used this cue directly to select D, whereas six first eliminated B and/or C as incorrect before making the association between "solid" and "volume." The remaining two students selected A and C, the latter after eliminating B as incorrect. All 15 test-naive students who used the stem-option cue selected option D, 13 directly and two following elimination of option B. Five (13.9%) test-wise students eliminated the first three options, as did one test-naive student. Two (5.6%) test-wise students indicated they knew the answer (one correct, the other wrong) and four (11.1%) reported they simply guessed. In contrast, seven (17.1%) students in the test-naive sample stated they knew the answer from a previous course (one correct, six incorrect) and 17 (41.5%) said they simply guessed (seven correctly). It is noteworthy that in contrast to the previous items discussed, the number of students in the test-naive sample who reported they guessed their answers to items 3 and 6 is greater than (and not equal to) that expected by chance alone. These results suggest that for these students something else other than chance was at work. When probed, many of these students were not able to articulate how they obtained their answer and concluded "I guess I guessed." It was though they had some sense of the similarity between the stem and option C (item 3) and D (item 6).

7. A major characteristic of natural resources is that they are
 A. sources of energy.
 B. nonrecyclable.
 C. nonrenewable.
 D. unevenly distributed. (Ministry of Education, 1986)

During the test construction phase, it was thought that students high in test-wiseness would recognize the similar options in item 7 and then select from among the remaining two-options and that students lower in test-wiseness would fail to see the similar options and therefore perform less well. However, the solution strategies adopted by students in both samples suggested that the material tested by item 7 was more familiar to the students than was the material tested in the previously discussed items. This seems likely given the frequent discussions of natural resources and their depletion in daily and weekly newspapers and on radio and television. In comparison to the previous seven items, the number of students in the test-naive sample who said they simply guessed their answer, two (4.9%), is significantly smaller, suggesting that they too possessed knowledge to answer this question. Indeed 27 (75.0%) students in the test-wise sample and 29 (70.7%) students in the test-naive sample avoided the two similar options. Eight test-wise students identified B and C as similar and used knowledge (4), deductive strategy ID1 (1), or guessing (3) to select their final answer. Only two test-naive students used similar options, one in conjunction with knowledge and the other in conjunction with ID1. Eight test-wise students employed ID1 to eliminate three options as incorrect. Seven test-naive students did likewise. Two test-wise and two test-naive students eliminated option (C). The two test-wise students then selected A (1) and D (1) using knowledge whereas the two test-naive students both guessed A. One test-wise student used stem-option similarity, citing the similarity between "resources" in the stem and "sources" in A, and another used "specific determiners" to eliminate B and C. Thirteen (36.1%) test-wise

and 28 (68.3%) test-naive students said they knew the answer. Finally, as indicated before, two (4.9%) test-naive students reported they guessed, as did three (8.3%) test-wise students.

Guessing

Following Millman (1966), the items designed to test guessing were scored as follows:

1. if a student selects an answer to an item where guessing is favored, score two. Otherwise score zero.
2. If a student does not select an answer to an item where guessing is not favored, score three. Otherwise score zero.

Given that six of the 10 items were guessing items and four were not, the total possible score was 24. The mean of the test-wise sample on the guessing subtest of the Interview Form was significantly greater than the mean of the test-naive sample (18.0 vs. 14.1; $p < .05$). Clearly, as expected from the performance of the students on the first section, the students in the test-wise sample were better able to identify and eliminate incorrect options on the four multiple-choice questions designed to encourage guessing.

Summary and Conclusions

Listed in Table 4 is a summary of the most frequently used strategies employed by the students in the test-wise sample and the students in the test-naive sample when they responded to the eight items included in Section 1 of the Interview Form, and the percentages of students in both samples who simply guessed their answers. Considered together, the greater use of test-wiseness strategies to answer an item by the test-wise sample and the marked reduction in guessing by the test-naive sample on item 7 with its familiar content illustrate the important role knowledge assumes in the application of the test-wiseness elements considered. Before students can apply a test-wiseness strategy to answer a multiple-choice item for which they do not know the answer and which, due to its flawed character, provides a test-wise cue, they must have some knowledge of the content being tested and that contained in the item's options. That the two samples compared differ in knowledge is supported by the observations of the principals of the schools in which students were interviewed. The principals pointed out that the students classified as test-wise were more academically talented, whereas the students classified a less test-wise or test-naive were among the less capable academic students. They further indicated that students in the test-wise sample were enrolled in a greater number of academic courses and were more likely to enter university than the students included in the test-naive sample.

The solution strategies used by the students also support the model of test-taking behavior presented in Figure 1. Clearly, if students did not have knowledge or believed they did not have knowledge relevant to the content tested, then, as illustrated by the following student responses, they were unable to engage a test-wise strategy in an attempt to answer the item:

Student A, Item 2:	Oh my gosh! (laughter) I guessed.
Student B, Item 5:	Guess this one too. I know nothing about this one too.

Table 4
Percentage of Use of Test-wiseness Elements

			Sample								
Test-wiseness element	*Item #*	*ID1*	*Test-wise (n=36)* *ID2*	*ID3*	*I IB4*	*G*[a]	*Test-naive (n=41)* *ID1*	*ID2*	*ID3*	*IIB4*	*G*
ID1	1	63.9	13.9	-	-	19.4	34.1	-	-	-	36.6
	8	80.6	-	-	5.6	2.7	61.0	-	-	7.3	19.5
ID2	4	8.3	58.3	-	-	25.0	-	12.2	-	-	61.0
ID3	2	8.3	-	55.6	-	13.9	2.4	-	-	14.6	53.6
	5	5.6	-	72.2	8.3	11.1	4.9	-	31.7	4.9	48.8
IIB4	3	36.1	-	-	44.4	22.2	12.2	-	-	31.7	43.9
	6	33.3	-	-	58.3	11.1	7.3	-	-	36.6	41.5
ID1, ID2	7	27.8	22.2	-	2.8	8.3	22.0	4.9	-	-	4.9

Note: Test-wise element counted whether final answer correct or incorrect and whether or not it was used alone or in combination with another.
[a]Simply guessed.

Student C, Item 4: Probably A ... I'm not sure, just guessing. I'm trying to eliminate answers out but I don't know ...

The strategy employed by students who possessed or believed they possessed relevant knowledge, but who did not find an answer using this knowledge alone, clearly followed the sequence shown in the dotted box of Figure 1. Typical of the responses of these students were:

Student D, Item 1: Not C because Stalin has to do with Russia ... A, I guess.

Student E, Item 1: Not C ... Stalin after the civil war. Um ... B and D both had to do with slaves and can't be right. So the answer is A.

Student E, Item 2: D is wrong! Oh, A and C are opposites so one is probably right.

(Probe: Why did you select C?)

I guessed between A and C.

Student F, Item 4: Cross out A and C ... they are the same.

(Probe: Why did you select D?)

It seems most complicated ... that's why.

Student D, after eliminating one option because it was known to him to be incorrect (element ID1), was unable to rule out other options. He then made a random guess. Student E first eliminated option C on item 1 using ID1. She then treated B and D as similar and pointed out they could not be correct (ID2). She then had a match and selected the test-wise response A. On Item 2, she again eliminated one option, D, as wrong (ID1). She then recognized that A and C were opposites and decided one was correct. Not having sufficient knowledge to decide, she guessed (ID3). Student F avoided options A and C on item

4 because they were similar (ID2). She then selected D on the basis of its complexity. The knowledge of these three students, although not sufficient to simply answer an item, coupled with their test-wiseness abilities, allowed them to produce a response after eliminating at least one option as incorrect.

The difference in approach observed between the test-wise and test-naive students can be explained by differences in cognitive monitoring. Although all six students quoted above initially experienced "cognitive failure" (Garner, 1990, p. 518), the test-wise students also experienced "metacognitive success" (p. 518). Unlike test-naive students who more frequently guessed their answers from among the options of an item they did not know, test-wise students were better able to take deliberate follow-up actions, integrating the partial content or substantive knowledge they possessed with their ability to take advantage of test-wiseness item cues or clues to systematically eliminate one or more of the options (Brown, 1987). Consequently the total scores of such students on a test containing flawed items will be spuriously inflated:

$$X_{tot} = X_{kd} + X_g + X_{twd/pk} + X_{er/pk} > X_{kd},$$

where

X_{tot} = the total number of correct responses.

X_{kd} = the number of correct knowledge derived responses, and the score a test was initially designed to illicit,

X_g = the number of correct random responses,

$X_{twd/pk}$ = the number of test-wise (integration of partial knowledge and test-wiseness strategies) derived responses, and

$X_{er/pk}$ = the number of "educated" random responses following application of test-wiseness.

Thus it appears that the effective application of test-wiseness reasoning strategies is dependent on some partial knowledge. This partial knowledge, although inadequate to respond to a test item solely on the basis of this knowledge, is sufficient when coupled with knowledge of the test-wiseness principles to increase the probability of correctly responding to items susceptible to test-wiseness. Students with low content knowledge but test-wise knowledge and students with partial knowledge but low test-wise knowledge will perform less well than students who possess both on such items (compare Garner, 1990, p. 520).

These results suggest that a test taker who possesses test-wiseness—partial knowledge about what an item measures and ability to take advantage of item cues—will have a greater probability of correctly answering a test-wise susceptible item than will a student low in test-wiseness. Thus a potential validity problem may exist when one attempts to interpret the meaning of a test score. If the partial knowledge component of test-wiseness is considered relevant when enhanced by a subject's capacity to utilize clues (flaws) present in items, then the interpretation of the total score as a valid indicator of achievement may be justified. But, to the extent this partial knowledge is considered not relevant, the interpretation of the test score will be confounded by construct-irrelevant easiness (Messick, 1989, p. 34). Further, if the intent is to compare

students based on their total scores, the difference between high and low scores on an examination containing test-wiseness clues will be larger than if only knowledge, or the lack thereof, was contributing to this difference. If test scores can be influenced by test-wiseness, then individuals involved in test development, administration, and/or interpretation need to carefully consider the construct of test-wiseness and how it affects scores.

Acknowledgments

The research was supported by the Social Sciences and Humanities Research Council, Canada. Reprinted with permission from the *Journal of Experimental Education, 59*(4), 331-350.

References

American Psychological Association. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Benson, J. (1988). The psychometric and cognitive aspects of test-wiseness: A review of the literature. In M.H. Kean (Ed.), *Test-wiseness* (pp. 1-25), Bloomington, IN: Phi Delta Kappa.

Brown, A.L. (1980). Metacognitive development and reading. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-481). Hillsdale, NJ: Erlbaum.

Brown, A.L. (1987). Metacognition, executive control, self-regulation and other even more mysterious mechanisms. In F.E. Wienert & R.H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Erlbaum.

Diamond, J.J., & Evans, W.J. (1972). An investigation of the cognitive correlates of test-wiseness. *Journal of Educational Measurement, 9,* 145-150.

Ebel, R.L., & Damrin, D.E. (1960). Tests and examinations. In C.W. Harris (Ed.), *Encyclopedia of Educational Research* (3rd ed.). New York: Macmillan.

Flavell, J.H. (1979, October). Metacognition and cognitive monitoring. *American Psychologist, 34,* 906-911.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research, 60*(4), 517-529.

Gibb, B.G. (1964). *Test-wiseness as a secondary cue response.* Unpublished doctoral dissertation, Stanford University. (University Microfilms, No. 64-7643)

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education and Macmillan.

Millman, J. (1966). *Test-wiseness in taking objective achievement and aptitude examinations.* Final Report, College Entrance Examination Board.

Millman, J., Bishop, H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25,* 707-726.

Ministry of Education. (1986). *Biology 12.* Victoria, BC: Author.

Rogers, W.T., & Bateson, D.J. (1991). The influence of test-wiseness upon performance of high school seniors on school leaving examinations. *Applied Measurement in Education, 4,* 159-183.

Sarnacki, R.E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research, 49,* 252-279.

Schuell, T.J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56,* 411-436.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. (1970). Grade level, sex, and selected aspects of test-wiseness. *Journal of Educational Measurement, 7,* 119-122.

Smith, J.K. (April, 1980). *The convergence strategy of test-wiseness.* Paper presented at the annual meeting of the American Educational Research Association, Boston.

Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356-442). Washington, DC: American Council on Education.

Thorndike, E.L. (1951). Reliability. In E.F. Lindquist (Ed.), *Educational measurement* (pp. 560-620). Washington, DC: American Council on Education.

Vernon, P. (1962). The determinants of reading comprehension. *Educational and Psychological Measurement, 22,* 269-286.

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 213-231*

Bikkar S. Randhawa
University of Saskatchewan

Theory, Research, and Assessment of Mathematical Problem Solving

This article describes theoretical and research issues in mathematical problem solving. Among the theoretical considerations are: nature of problem solving, problem solving approaches, situated and context-independent cognition, externalization of internal representation and its role in the emergence of genius solutions to enduring problems, and the development of competence. The research issues identified concern the assessment of mathematical problem solving of students so that meaningful and contextually relevant analysis of the problem solving process can take place. A study is described in which this approach is exemplified. Also, differences in cognitive and metacognitive processes of high and low mathematics ability students in solving five everyday mathematics problems are related to the literature on the development of expertise.

Cet article décrit des questions théoriques et des questions de recherche dans le domaine de la résolution de problèmes en mathématiques. Parmi les considérations théoriques on inclus: la nature de la résolution de problèmes, des approches à la résolution de problèmes, la cognition située et indépendante des contextes, l'extériorisation de la représentation interne et son rôle dans l'émergence de solutions géniales aux problèmes persistants, et le développement de la compétence. Les questions de recherche identifiées traitent de l'évaluation en résolution de problèmes en mathématiques des élèves de façon à ce que l'analyse des processus de la résolution de problèmes soit significative et vue dans son propre contexte. Une étude illustrant cette approche est présentée. En plus, les différences entre les processus cognitifs et métacognitifs des élèves d'habiletés supérieures et inférieures en mathématiques à résoudre cinq problèmes mathématiques ordinaires sont interprétées selon la littérature dans le développement de l'expertise à la résolution de problèmes.

This article deals with theoretical and research issues in mathematical problem solving. Theoretical issues concern the nature of problem solving, approaches to problem solving, situated and context-independent cognition, externalization of internal representation, and the development of competence. The research and assessment section deals with one of our research studies in mathematical problem solving and gender differences, not a primary focus here.

Problem solving in general and in specific mathematics domains have long occupied researchers. This area received more impetus with the reemergence of cognitive psychology in the 1950s. It is not surprising that this interest coincided with an increased interest in including critical and creative thinking curricula in grade schools and postsecondary institutions. Often critical and creative thinking was seen as a new discovery. Its advocates began promoting the development of these abilities as if past generations had grown up without

Bikkar Randhawa is a professor of educational psychology with current research interests in assessment, cognitive psychology, and gender-related differences in learning, instruction, and problem solving in mathematics.

them. Ironically, those promoting these abilities were themselves schooled where these abilities were not emphasized. The question arises how and why those who began to see the need to develop problem solving competence came to this realization. Other related questions are: How do people acquire problem solving competence? Are those who receive direct or curriculum-infused instruction in problem solving competent in domain-specific contexts or in general situations? In other words, do these abilities transfer in nonsituated applications?

Mathematics is a subject in which problem solving is most critical (Polya, 1962). Despite the importance of problem solving ability, however, little is known about how to teach people to solve problems. A great mystery of our time is why some people solve problems for which they have all the necessary component skills and others fail.

Polya's (1962) contribution in the domain of mathematical problem solving received general acclaim (Scandura, Durnin, & Wulfeck, 1977). Also, there was intrinsic support for his notion of heuristics. "Although often useful, his heuristics are frequently little more than general hints, and leave much to be desired insofar as pinpointing what a human or computer must know in order to solve specific kinds of problems" (1977, p. 166). Theoretical and research issues in mathematical problem solving, as evidenced above, are complex. Complexity does not mean we should ignore or not debate these issues. This article provides a brief discussion of the issues identified above.

Theoretical Issues

The Nature of Problem Solving

Problem solving has meant different things to different people and has given rise to different interpretations in the same or different domains of study. Problem solving is not restricted to mathematics. Creating new ideas, for example, is a form of problem solving. Troubleshooting and inventing new products or techniques are other forms of problem solving. Branca (1980) states:

> Although problem solving in mathematics is more specific, it is still open to different interpretations. Activities classified as problem solving in mathematics include solving simple word problems that appear in standard textbooks, solving nonroutine problems or puzzles, applying mathematics to problems of the "real" world, and creating and testing mathematical conjectures that may lead to new fields of study. (p. 3)

Problem solving has meant (a) a goal, (b) a process, and (c) a basic skill. This characterization does not include the role of knowledge. Greeno (1980) emphasized the important role of knowledge in problem solving. Knowledge of a discipline is important in problem solving in that discipline and that is precisely what we know from research on expertise and problem solving. Greeno described the distinction that used to be made between performance based on knowledge, that is, applying an algorithm or just following a remembered set of sequences, and performance based on problem solving. The latter was judged more highly than the former. Here the value on performance was placed by the judge. If the performance was such that the judge could identify the representative knowledge used in solving the problem then the perfor-

mance was judged less highly than if the knowledge base for the performance was not identifiable by the judge. Because of the research in problem solving this distinction is much eroded. But folklore still prevails. Greeno (1980) argues:

> All problem solving is based on knowledge. A person may not have learned exactly what to do in a specific situation, but whatever the person can do requires some knowledge, even if that knowledge may be in the form of general strategies for analyzing situations and attempting solutions.
> The other reason that the distinction between knowledge-based performance and problem solving has eroded is that we now can characterize the performance of individuals who solve problems; and when we carefully consider the performance that occurs in more routine situations, we find that the essential characteristics of real problem solving are there also. (p. 10)

Often major components of problem solving are present in situations where we do not regard the performance as an instance of problem solving. So it is important to distinguish between: (a) situations where the problem solver has specific domain knowledge that considerably facilitates problem solving; and (b) situations where the problem solver must use more general knowledge and procedures to solve a given problem. But the specificity of relevant knowledge is a matter of degree, not kind (Greeno, 1980). Problem solving performance should be considered a continuum, not a dichotomy.

The value we place on problem solving performance has implications for both instruction and assessment. If we do not value the role of specific domain knowledge in problem solving, it implies that instruction in the specific underlying knowledge would not be necessary. Therefore, instruction will emphasize only generic problem solving skills and strategies so that performance will merit high value when it is judged. Proportional weights for domain-specific knowledge and generic problem solving skills and strategies would be required in order to undertake meaningful assessment in this area.

Approaches to Problem Solving: Representational and Strategic Processes

Mathematical problem solving task can be broken down into two major component processes: (a) problem comprehension, and (b) problem solution (Mayer, 1985, 1986; Mayer, Larkin, & Kadane, 1984). Problem comprehension involves translation of each sentence in the problem into an internal representation and integration of the literal information to form a coherent structure. Problem solution, on the other hand, requires planning, monitoring, and executing the requisite mathematical operations. Gagne (1983), however, suggests that there are three phases in the performance of a mathematical task: translation, execution (problem solution), and validation or monitoring of the problem solution. It has been found that students usually have trouble in the problem comprehension or translation phase (Mayer, 1985, 1986). However, most mathematics instruction focuses on the problem solution phase, particularly algorithmic computations (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1981; Randhawa & Beamer, 1990). Also, students put little emphasis on the monitoring phase.

In somewhat the same vein Kintsch and Greeno (1985) presented a processing model that dealt explicitly with both the text comprehension and problem

solving aspects of arithmetic word problems. The main features of this model include a set of knowledge structures and strategies for using them in building a representation and solving a problem. The verbal input is transformed into a conceptual representation of its meaning that takes the form of a list of propositions. The knowledge structures of this model comprise (a) a set of propositional frames, (b) a set of schemata that represent properties and relations of sets in general form, and (c) a set of schemata that represent counting and arithmetic operations in general form.

Kintsch and Greeno (1985) presented evidence to suggest that the general features of the comprehension process are alike in a variety of situations such as mathematical word problems, reading a story, and reading a stock market bulletin, but that the content of the comprehension strategies, the nature of the specific knowledge structures, and the type of resulting macro structures are task- and goal-specific. In their view, comprehension of a mathematical word problem is achieved by constructing a conceptual schema from the verbal form of the problem on which problem solving processes can operate. For example, understanding a mathematical word problem would constitute conceptual relations among numerical values that guide the selection of calculations to be done. Students' experience in solving word problems "results in their acquiring a special set of strategies for constructing mental representations of texts that are suitable for applying mathematical operations" (p. 110). In this sense, understanding and solving word problems can be regarded as "expert" performance as cited in the literature on problem solving in physics (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980).

Sweller, Mawer, and Ward (1983) found that mathematics experts (graduates) and novices (9-12-year-olds) employed distinctive strategies while solving mathematical problems. Expert problem solvers preferred forward-chaining strategies, using a large formula, as opposed to various steps one at a time, whereas novices typically employed a means-ends approach, working backward from the target solution. The novices' use of means-ends strategies delayed acquisition of an appropriate schema. Also, the use of nonspecific rather than specific goals was found to facilitate acquisition of expertise. The course of development of expertise witnessed the switch from the means-ends strategy to the forward-chaining strategy. Also, as expertise was acquired, subjects showed the conventional concomitants of expertise such as a reduction in the number of moves required for solution. These results and their conclusion flout conventional wisdom in presenting problem structure. For instance, they conclude

> that the real goal of problem solution needs to be analyzed. If the goal is simply to solve a particular problem, then a conventional problem format is satisfactory. Alternatively, a reduced goal-specificity procedure might be preferred if, as normally occurs in educational contexts, the real goal is to gain additional insights into the problem structure and the structure of the subject matter that gave rise to the problem. (p. 661)

Hall, Kibler, Wenger, and Truxaw (1989) analyzed the episodic structure of written protocols of 85 undergraduates who were asked to solve four story problems. Their results showed that comprehension and solution of the problems were complementary activities, rather than distinct phases of a prob-

lem solving task as Gagne (1983) and Mayer (1985, 1986) proposed. Also, Hall et al. (1989) found that the two complementary activities in the problem solving process resulted in a succession of episodes. Furthermore, competent problem solvers used various forms of model-based reasoning to identify, pursue, and verify quantitative constraints required for the solution. Russell and Ginsberg (1984) reported that fourth graders with mathematical difficulties or novices lacked sophisticated understanding and strategies, but they were not seriously deficient in basic mathematical concepts and nonalgorithmic procedures.

Problem solvers' earlier experiences with a genre of problems are linked to their self-confidence, a crucial factor in problem solving (Lester, 1980). Self-efficacy, confidence in one's ability to do a task rather than actually doing it, is important for cognitive functioning (Bandura, 1986; Davis, 1973). Furthermore, Bandura (1986) and Polya (1957) emphasized the role of the problem solver's interest and motivation. They also stressed that those presenting the problem for solution must choose an appropriate level of difficulty and present it in an interesting way.

Snyder (1987) identified two types of people in social contexts, high self-monitoring and low self-monitoring. High self-monitors tend to be sensitive to feedback and use this to provide cues to guide their behavior. Low self-monitors are not as responsive to the environmental cues. Wong (1988) tested this theoretical formulation in mathematics problem solving and found that high self-monitoring individuals were significantly better at applying rules and in general problem solving abilities.

Lewis and Mayer (1987), in a study involving two experiments, investigated college-level students' difficulties in comprehending relational statements in arithmetic word problems. Their results supported the contention that students' errors on many mathematical problems result from difficulty in the comprehension phase rather than in the solution phase. The proposed theoretical model states that "the problem solver comes to the problem solving task with a set of schemata or preferences for the form of assignment and relation statements in compare problems" (p. 370). When the form of the relational sentence is incongruent with the problem solver's schema, he or she must restructure it, and this rearrangement or transformation may lead to representational errors. This formulation for the specific type of mathematical problem can be deduced from the general schema preference or bias of the problem solver. For problems that are not similar in structure, the general expectation for each is derived from earlier experience (Bandura, 1986).

Situated and Context-Independent Cognition

People bring their personal schemata to problem solving tasks. These personal schemata, knowledge structure, strategies, and preferences, guide how people approach problem solving. Critical to developing problem solving schemata is experience with the genre of problems (Kintsch & Greeno, 1985; Sweller et al., 1983). It is reported that boys have more access to problem solving experiences outside the mathematics classroom than do girls (Kimball, 1989; Hyde, Fennema, & Lamon, 1990), and it is possible this experiential advantage would translate into differential expertise of boys and girls on similar tasks or tasks that require situated learning (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Because of this experiential advantage outside the classroom,

even when boys and girls are matched for academic achievement in mathematics, boys' cognitive and affective schemas might be more appropriate for handling everyday mathematical word problems.

Knowledge is situated because it is in part a product of the activity, context, and culture where it is developed and used. This view of knowledge affects our understanding of learning and problem solving. Brown et al. (1989) argue that conventional schooling too often ignores the influence of school culture on learning and problem solving. The Cognition and Technology Group at Vanderbilt (1990) propose anchored instruction to overcome the problem of *inert knowledge* (Whitehead, 1929), which is knowledge that can usually be recalled when individuals are explicitly asked to do so but is not used spontaneously for problem solving even when relevant. Situated cognition and anchored instruction, which provides a way to recreate some advantages of apprenticeship training in formal educational settings involving groups of students, suggest ways to think differently about instruction for and assessment of problem solving.

Externalization of Internal Representation

Problem solving in any domain requires, among other things, encoding and translation, which are internal and unobservable. To understand the process of problem solving these internal mechanisms must be externalized. The form of these representations is not always verbal; sometimes internal representation consists of internal perceptual images. Visualization as a strategy in problem solving and skillful performance is assumed to play an important role in many problem solving situations (Paivio, 1978; Shepard, 1978). Whether verbal or nonverbal, externalization of internal images offers a richer data base for understanding problem solving processes. Shepard (1978), in particular, depicts a large array of his externalization of internal images. He also documents the use of imagery by many well-known creative geniuses including Albert Einstein, Hermann von Helmholtz, Sir Francis Galton, James Clerk Maxwell, and James D. Watson.

As shown above, visualization was important for many important discoveries. Therefore, it must have implications for mathematical problem solving in a variety of contexts. But problem solving is discovering new applications of acquired knowledge (Greeno, 1980), and "knowledge itself comes to form a part of a person's context of problem solving" (Sternberg & Salter, 1982, p. 19). The challenge is to seek ways to assess the visualization. Assessment of visualization is as challenging as assessment of other internal thought processes engaged in by the problem solver during the process of problem solving. In this context it is imperative to consider the development of competence in problem solving.

Development of Problem-Solving Competence

Competence is the capacity to perform to a set of criteria for a job, task, or ability. It underlies the accumulation, through experience and training, of the required domain knowledge base to function adequately when given an opportunity to perform. The observable expression of competence—performance—is viewed in the literature as depending in part on the context in which performance occurs (Zigler & Seitz, 1982).

Competence in problem solving develops with experience in a situated context (Brown et al., 1989; Lave & Wenger, 1991). Problem solving competence in a cognitive domain is much like a skill in a particular domain in that its transfer is limited. Zigler and Seitz (1982) give an example of a handicapped individual who could operate a complicated piece of machinery competently; but when the machine was rotated 30 degrees, the individual was confused and unable to operate it. One could argue that this person was less competent than a skilled machine operator; more to the point, however, is the fact that, in a well circumscribed context, he or she was able to demonstrate skilled performance. What debilitated this individual's performance was its unfortunate lack of cross-situational transfer.

Sternberg and Salter (1982) point out important assessment-related issues in this area. They state:

> The sensitivity of performance to context suggests that methods to assess competence—that is, ways of attempting to elicit performance—must be designed with consideration of contextual factors, and must have considerable diversity to reflect the diversity of possible contexts in which performance may occur. This observation obviously has implications for testing and measurement. It also seems to require that a full theory of intelligence address the ways in which various contexts may affect the expression of a given competence. (p. 18)

Because developing problem solving competence depends on experience in a specific context, the level of competence at various ages would vary. Thus the range of individual differences in the expression of competence in mathematical problem solving should parallel that in other cognitive abilities. This experiential perspective of development of competence in mathematical problem solving does not exclude genetic factors; rather, it regards these sources of individual differences as putting an upper limit to the extent to which competence can be developed.

Research and Assessment of Mathematical Problem Solving

This section describes a research study in which assessment of mathematical problem solving was undertaken. This example should underline an approach to authentic assessment of problem solving in which multisource data were recorded for analysis and evaluation.

Introduction

Knowledge, transformational operations, and concomitant skills or models are necessary but not sufficient for skilled performance. At times people do not perform optimally although they may possess all the necessary knowledge structures and strategies. This is because "self-referent thought mediates the relationship between knowledge and action" (Bandura, 1986, p. 390). The representation of the past experiences involves constructive processes, which personalize the representation (Bandura, 1986). Moreover, environmental events are filtered through personal meanings and biases and are cognitively transformed into propositional knowledge. When subjective environmental experiences are not congruent with belief, memory for such past experiences is distorted (Cordua, McGraw, & Drabman, 1979; Signorella & Liben, 1984).

The purpose of this study was to assess the cognitive and metacognitive skills, problem solving achievement, and related processes of high and low mathematics ability students on mathematics problems.

Method

Subjects

Forty subjects who participated in this study were selected from a population of 225 grade 12 students in nine classes from three high schools in a medium-sized city in midwestern Canada. All 225 students were enrolled in Algebra 30, an academic grade 12 algebra course, and all had taken the Mathematics Ability Test (MAT) at the beginning of their Algebra 30 course. The MAT is a 40-item test representing the content of academic grade 11 mathematics (Randhawa, Ng, & Beamer, 1990).

Twenty students, 10 boys and 10 girls, were selected from the top quartile of the MAT test (a high ability or HA group); and another 20 students, also 10 boys and 10 girls, were selected from the bottom quartile of the MAT test (a low-ability or LA group). Students were matched pairwise by gender.

Instruments

Development of everyday mathematical problems. Twenty-eight mathematical word problems reflecting real-world, everyday situations were developed. These problems were considered appropriate for those students who had finished the grade 11 academic mathematics course. The problems covered a variety of mathematical content areas including area, proportion, ratio, graphing, tables, charts, percentages, multiplication, and division. The situational content of the problems entailed, for example, car rental, sale prices, currency exchange, and discounts.

Two grade 12 teachers were enlisted, along with two of the researchers, to rate the 28 problems on their relevance to everyday life and their level of difficulty. Because of the ratings only 24 problems were retained for use in the study. Three problems rated as easy were used as warm-up problems, and every student attempted one of these at the beginning of the first videotaping session. Five were selected as the core problems and every student was required to solve these. The problems appear in Appendix A.

The remaining 16 problems were designated as optional and were administered to the subjects only to get provisional data for assessing their effectiveness. With three exceptions, data for only eight subjects for each of these optional problems could be collected under the time constraints of the study. For three problems data were collected on fewer than eight subjects. Therefore, other details on the classification and administration of the optional problems are not relevant for the purposes of this study.

Each problem was prepared on a single sheet, which allowed for work space for students to use during their problem solving session. Additional paper was provided as needed. Some problems had documentary attachments on separate sheets.

Procedure

Each student attended two sessions. In the first each student was given nine problems to solve. In order of presentation, each student was given a practice problem, an optional problem, five core problems in the same order to each

subject, and two of the optional problems. The students were asked to think aloud while solving the problems as it was a purpose of the study to find each subject's representation of the problem as clearly as possible. The subject worked on the solution of the problem uninterrupted by the experimenter unless the subject stopped verbalizing the procedures and strategies being executed. In that case the experimenter would encourage the subject to reveal his or her mental operations by verbalizing.

Each session was videotaped. For this purpose an Elmo Visual Presenter, EV-308, was interfaced with a VCR and a monitor. Through a camera focused on the problem sheet, this equipment can capture the written work, and through a microphone the oral record of the subject; confidentiality was maintained as the subjects themselves were not pictured.

In the second session each subject was interviewed and questioned on the way he or she had represented each problem, the strategies used, and the various procedural steps taken to solve the problem. During this interview the subject was shown his or her work on the problem sheet from the first session. The equipment used in the first session was again used in the second session for a written and oral record of the interview.

One of the two researchers interviewed the subject. The other researcher was present at all the sessions to operate the videotape machine and to take additional notes when required.

After each subject had been interviewed on the specifics of each problem, he or she was asked to rate the problem on its familiarity (how familiar he or she was with the specific kind of problem) and everyday occurrence (what was the perception of the subject about how often a specific type of problem was encountered by people in their everyday business) using a 5-point scale. Only the end points of the familiarity scale were described: 1 = "not at all familiar"; and 5 = "extremely familiar." Descriptors for the everyday occurrence scale were: 1 = "never"; 2 = "once or twice a year"; 3 = "about once a month"; 4 = "about once a week"; and 5 = "everyday."

Each session took about an hour. The interview timetable for each student was arranged such that the second session was on the school day after the first session. The subjects were paid for their participation in the experiment.

Analysis

Problem Solving Skills

Videotaped segments of each respondent, while solving the five core problems and while being interviewed about the same five problems, were analyzed by two trained coders. The coders used a five-point Likert scale to rate the level of evidence of nine cognitive and metacognitive skills of each subject. The five points of this scale were: 1 = "no evidence"; 2 = "below average evidence"; 3 = "average amount of evidence"; 4 = "above average evidence"; and 5 = "substantial evidence." For each subject the ratings of each skill evaluated by the two coders were combined. The consistency between the two coders for rating each skill was determined by computing the Pearson correlation coefficient.

The nine cognitive and metacognitive skills rated were: declarative knowledge; procedural knowledge; translation of mathematical information; use of heuristics; use of estimation or checking or monitoring the solution; level of confidence; ability to transfer the current knowledge; ability to general-

ize/abstract; and mental computing facility. These categories are often mentioned in the literature as important problem solving skills (Bandura, 1986; Carpenter et al., 1981; Chi et al., 1981; Mayer, 1986), but assessment of these skills using rating scales was adapted from a recent study (Randhawa & Beamer, 1990).

The sum of the ratings of the two coders was analyzed in a 2 x 2 (Ability Level x Gender) fixed factorial analysis of variance (ANOVA) design for each cognitive and metacognitive skill. Students' ratings of their familiarity with each problem and of their perception of the occurrence of the kind of problem in everyday transactions were analyzed for each problem in the same ANOVA design.

Accuracy and Monitoring

Significance of differences of percentages of students who monitored their work/solution for each problem for the two ability levels and gender categories was tested.

Each core problem was scored by two independent scorers according to a predetermined scoring scheme. This scheme took into account both the process and computational steps completed for each problem. Depending on this analysis, the total score for a problem ranged from 4 to 17. Score assignments for the process and computational components of each problem are provided in Appendix A.

Using the 2 x 2 ANOVA design described above the total score on the five core problems and the scores on each problem were analyzed. Despite the repeated measurement of the five problems, it was felt that analysis of each individual problem would provide more insight into the differences for the competence level and gender. A further justification is that each problem requires different strategies and representational processes for solution.

Table 1
Mean of Combined Ratings of Cognitive and Metacognitive Skills by Ability and Gender

	Ability		*Gender*		MS_e
Problem Solving Skill	*High*	*Low*	*Male*	*Female*	
Cognitive					
Declarative Knowledge	7.75*	4.35	6.50	5.60	3.68
Procedural Knowledge	8.25*	4.85	7.15*	5.95	2.58
Translation of Information	7.45*	4.35	6.70*	5.10	2.47
Ability to Transfer Current Knowledge	7.35*	4.40	6.20	5.55	2.30
Ability to Generalize/Abstract	6.95*	4.65	6.05	5.55	2.26
Computing Facility (Mental)	7.75*	5.35	5.70*	6.40	1.03
Metacognitive					
Heuristics	7.40*	4.65	6.50	5.55	2.90
Estimating/Checking	7.30*	4.75	6.50	5.55	2.86
Level of Confidence	7.50*	4.20	6.60*	5.10	2.54

*$p<0.05$.

Results

Cognitive and Metacognitive Skills

Consistency estimates of the two raters' analysis of various cognitive and metacognitive skills of students were in the range .904 to .956.

Means of combined ratings of cognitive and metacognitive skills by ability and gender are provided in Table 1. Also shown in this table are the MSe for the univariate effects. The profile of means of the HA group was significantly different from that of the LA group, multivariate $F(9, 28)=7.06$, $p<.05$. The ANOVA F-ratios with 1 and 36 degrees of freedom indicated that the ability main effects were significant ($p<.05$) for all the skills investigated. In each case, as expected, the high achiever group (HA) were rated significantly higher than the low achiever group (LA).

The multivariate gender main effect was significant, $F(9, 28)=3.45$, $p<.05$. But the univariate Gender main effects were significant ($p<.05$) only for the Procedural Knowledge, Translation, Confidence, and Mental Computing. In the Use of Heuristics and for the Ability to Estimate the Answer, for the gender effect, the F-ratios only approached significance ($p<.10$). It appears that boys were more skillful in Procedural Knowledge and Translation, and approached mathematical problem solving tasks with more confidence than their female counterparts. Girls, on the other hand, were significantly better than their male counterparts in Mental Computing. On the remaining five skills, the observed means of boys were higher than those of girls, but the differences were not statistically significant. The multivariate Ability x Gender interaction was not significant.

Frequency of Occurrence of the Core Problems

Means of students' ratings of the frequency of occurrence (encounter) in everyday life of the five mathematical problems they solved are given by ability and gender in Table 2. Generally, the HA students indicated more favorably than the LA students that each of the five mathematical problems occurred frequently in daily commerce and life. But significant F-ratios ($p< .05$) were obtained only for problems 3 and 4.

Table 2
Means of Student Ratings of the Core Everyday Mathematics Problems for their Frequency of Occurrence by Ability and Gender

	Ability		Gender	
Math Problem	*High*	*Low*	*Male*	*Female*
Core Problem 1	2.30	1.80	2.20	1.90
Core Problem 2	3.65	3.10	3.55	3.20
Core Problem 3	2.40*	1.60	2.25	1.75
Core Problem 4	2.70*	1.65	2.60*	1.75
Core Problem 5	2.45	1.85	2.55*	1.75

*$p< 0.05$.

Table 3
Means of Student Ratings of the Everyday Core Mathematics Problems for their Familiarity by Ability and Gender

Math Problem	Ability		Gender	
	High	Low	Male	Female
Core Problem 1	3.10*	2.35	2.90	2.55
Core Problem 2	4.15	3.65	3.80	4.00
Core Problem 3	3.25*	2.20	2.85	2.60
Core Problem 4	3.35	2.60	3.20	2.75
Core Problem 5	3.50*	2.70	3.15	3.05

*$p < 0.05$.

Similarly, boys generally rated the frequency of occurrence of the five problems higher than did girls, but only for problems 4 and 5 the *F*-ratios were significant at the .05 level. Although these problems were intended to be everyday problems and the judges felt them to be so, the students did not perceive them as everyday mathematical problems.

Familiarity Ratings

Means of students' perceived familiarity with each problem by ability and gender are given in Table 3. Competent students' means of ratings on problems 1, 3, and 5 were significantly higher ($p<.05$) than those of LA students. On the remaining two problems the means of the HA group were higher than those of the LA group, but the differences were not significant.

Table 4
Percentages of Correct Answers on Five Core Mathematics Problems by Ability, Gender, and the Total Group and Standard Errors (s_e) for Testing the Significance of Two Independent Proportions

Problem	Ability Level		Gender		Total	s_e of p
	High (n =20)	Low (n =20)	Male (n = 20)	Female (n = 20)		
1. Part 1	80*	20	65	35	50.0	.158
1. Part 2	45*	5	30	20	25.0	.137
1. Part 3	80*	10	55	35	45.0	.157
1. (combined)	68*	12[a]	50	30	40.0	.155
2.	60	30	55	35	45.0	.157
3.	60*	20	55	25	40.0	.155
4.	50*	15	45	20	32.5	.148
5.	30	15	35	10	22.5	.132
Total	54*	18	48	24	36.0	.151

[a]These are the average percentages obtained by averaging those of the three parts.
*$p < .05$

There were no significant differences between the means of boys and girls on the familiarity ratings of any of the five core problems.

Correct Final Answers

Percentages of respondents getting the correct final answer on each problem by ability and gender are given in Table 4. Also, this table gives percentages of the total group who obtained the correct answer for each problem or subproblem. An examination of these percentages allows a quick conventional analysis of the difficulty of obtaining the correct answer for each problem. As can be seen, in terms of the correct final answer obtained, problem 5 was the most difficult (22.5%) and problem 1.1 was the easiest (50%).

The HA group, as expected, obtained a significantly higher proportion of correct answers on each problem, except problems 2 and 5, than the LA group. There were no significant gender differences in the proportions of correct answers.

Monitoring (Checking) Work/Answers

Table 5 shows the percentages of those who monitored or checked the work during each problem solving session by Ability x Gender and the total group. Differences between the ability and gender categories for the independent proportions on each problem were not statistically significant. This suggests that competence level or gender did not affect the monitoring activity.

Accuracy Scores

Means of accuracy scores for the five core problems and mean square errors are shown in Table 6, which also shows corresponding values for the combined accuracy scores of the five problems. The multivariate effect for testing the differences between the mean vectors of the two ability groups was significant, $F(5, 32)=6.80$, $p<.05$. The ANOVA results showed that the ability main effect was significant ($p<.05$) for all except problem 2. As shown in Table 6, the HA group had higher means on all the problems than the LA group. The same

Table 5
Percentages of Students Who Checked (Monitored) Their Work/Answers by Ability, Gender, and the Total Group

Problem	*Ability Level*				*Total*
	High		*Low*		
	Male (n = 10)	*Female (n = 10)*	*Male (n = 10)*	*Female (n = 10)*	
1. Part 1	20	60	60	20	40.0
1. Part 2	20	10	10	10	12.5
1. Part 3.	30	40	0	20	22.5
2.	10	20	0	30	15.0
3.	20	50	60	80	52.5
4.	40	50	60	30	45.0
5.	20	30	20	40	27.5

Note: There were no significant differences between the Ability and Gender categories.

Table 6
Means of Accuracy Scores on the Core Everyday Mathematics Problems by Ability and Gender

	Ability		Gender		MS_e
Math Problem	High	Low	Male	Female	
1. (9)[a]	5.70*	1.77	4.20	3.27	5.27
2. (8)	5.70	3.95	5.20	4.45	9.83
3. (5)	3.70*	2.20	3.45*	2.45	2.39
4. (4)	1.80*	0.70	1.87	0.62	2.83
5. (17)	12.60*	8.40	10.75	10.25	13.67
Total (43)	29.50*	17.02	25.47	21.05	63.79

[a]These are the possible scores.
*$p < 0.05$.

result, obviously, was obtained for the total accuracy scores on the five problems.

An examination of Table 6 suggests that boys had higher accuracy score means than girls on the five problems, but the multivariate gender main effect for the five accuracy scores on the core problems was not significant. Also, the multivariate Ability x Gender interaction was not significant.

Discussion

The results of this study suggest that competence is characterized by a rich knowledge base and metacognitive strategies. Competent individuals exhibit, much like experts, a confident approach to solving problems (Lester, Garofalo, & Kroll, 1989) that is the result of high perceived self efficacy (Bandura, 1986). Not only did the HA individuals work more accurately and efficiently (Chi et al., 1981; Glaser, 1989), they also found the problems presented familiar to them and perceived problems such as these are to be encountered somewhat frequently in everyday transactions outside the classroom. The procedural skills of the HA subjects were superior to those of the LA individuals, and they also obtained a higher percentage of correct answers than the LA group. It appears that the structure of knowledge of those with high mathematical ability, whom we characterize in this study as HA performers, was pervasive in the solution of the everyday problems (Chi et al., 1981; Feigenbaum, 1989). Also, as suggested by Ericcson and Staszewski (1989), it might be because of superior knowledge structures and cognitive and metacognitive skills that our HA subjects exhibited significantly higher levels of efficiency, judgment, heuristics, and extrapolation in solving mathematical problems than the LA subjects.

From the oral and written protocols of the subjects, it was obvious that the HA group had acquired automaticity in the use of algorithms, and their approach to problem solving was governed by conditions in the problem statement. The LA group, on the other hand, was unsystematic in the use of mathematical operations and at times applied various operations by trial and error to obtain a reasonable answer. Interestingly, however, the LA group, much like the HA group, had an idea of an answer that would be reasonable

but did not know how to arrive at it; that is, they lacked procedural skills (Chi et al., 1981). For example, when working on problem 1.1, which required direct application of a percentage shown in a pie graph to a given population, many students from the LA group multiplied the given number by the percentage figure and obtained a much larger number than the total population. When they realized the answer was unreasonable, they divided the total population by the percentage. After checking the answer they realized again that it was unreasonable. Then, sometimes, they would read the problem again and inspect the pie graph and convert the percentage to its equivalent proportion. Again, trial and error governed their behavior, and even after they had finished exploring options and picked a solution, they were not certain that they had obtained the correct solution.

The total accuracy by problem of the HA group was better than that of the LA group. Thus it may be concluded that the cognitive schemata for mathematical problem solving of HA subjects have the contrasting properties outlined above. Instruction in mathematical problem solving might be more relevant and effective if it included both cognitive and metacognitive processes (Glaser, 1989); if it were recognized that expertise is acquired gradually; and if success experiences were recognized as frequent affective ingredients in mathematical problem solving and in building the new cognitive architecture (McLeod, 1989).

Although it was not the major thrust of this article, unlike much other research on everyday problem solving, this study investigated gender differences in cognitive and metacognitive processes. Consistent with Kimball's (1989) synthesis, our results showed that in spite of equal exposure to formal mathematical instruction and curriculum, differential outside experiences of men and women might produce processing differences. The reported differences in boys and girls who were matched on total mathematics achievement were observed on certain cognitive and metacognitive processes. Girls showed higher mental computing facility than boys, but boys were significantly higher on procedural knowledge, translation of problem information, heuristics, and level of confidence. These results would suggest that men and women have acquired different cognitive structures for processing mathematical problems. Similar differences in processing mathematical instruction in natural classroom environments in grade 4 were reported in a recent study (Randhawa & Beamer, 1990). Taken together, the results of these studies would lead us to speculate that processing differences in boys and girls begin at even earlier age levels. Socialization and sex stereotyping phenomena have been ruled out by some in western cultures as plausible explanations. Then what other factors might be instrumental in producing the observed differences? Perhaps the old interactional resolution to the nature and nurture controversy might be applicable here. That is, biogenetic differences and societal interventions jointly produce enduring differences (Kimball, 1989). Differences ought not be considered as something that must be eradicated, but as something we should capitalize on to use effectively what individuals and groups have to offer. Meanwhile, instructional innovations should be explored to optimize the potential of boys and girls in mathematics problem solving so as to develop a reasonable degree of expertise.

Research on humans and computers seems to suggest that one way to change a novice problem solver into an expert is to provide many examples—including problems that are worked out by the instructor (or book) and problems that are worked out by the student. The examples must help the novice to see the step-by-step procedure, rather than just the final answer (Sweller et al., 1983). Just such an approach, process of problem solving as opposed to the product of problem solving, was first used by Bloom and Broder (1950) to help remedial students succeed in comprehensive examinations at the University of Chicago. Such an approach seems to contradict the idea that general thinking skills can be easily taught; in contrast, it shows that thinking skills and strategies are learned in a particular context such as algebra or geometry and that becoming an expert problem solver takes a lot of specific learning. For example, in our study most high ability students as compared with low ability students used one step calculation in Problem 2 to determine the discounted price of the item in each store, that is, the student would say that at Store A the price will be 20% off, which is 80% and at Store B the sale price is 10% off which has already been reduced by 15%. So Store B would charge 90% of 85%, 76.5%. Thus Store B has a better price. These students demonstrated forward-chaining in such procedural large-step calculations. In contrast, most low ability students would use the calculator to calculate 10% of an assumed price, then would subtract this amount from the original. Similarly, to calculate the price at Store B, they would take several discrete steps, sometimes in a haphazard fashion, and then might or might not eventually compute the correct price. The low ability students thus took longer to find the answer and used means-ends strategy in such situations. Problem 1 similarly posed difficulty for these low ability students, whereas high ability students were generally successful in representing the problem properly and executing the solution by undertaking generally bigger steps toward the solution. Also, the high ability students had more confidence in their answer. The progressive differentiation of students by ability in a domain reflects differences in their development of expertise. Unless remedial and specific training is provided in the domain of interest, marked differences in expertise will emerge. Training would reduce but not eliminate individual and group differences in expertise.

The research example provided here goes beyond understanding differences in problem solving expertise. It illustrates how the problem solving process can be captured and assessed. Not only is it important that think-aloud protocols be rendered as reliably as possible, but that the actual process of problem solution be juxtaposed with it so that analysis and subsequent clinical interview are meaningful and contextually relevant. Our use of the specialized recording device made this possible. For the assessment of cognition it is important to pay special attention to the externalization of problem solvers' internal schemata. Thus videorecording of think-aloud protocol along with the accompanying solution process provides initial data. These data ought to be enriched, as in this study, through clinical interviews to maximally capture the cognitive structure of the problem solver. Having data in such form can provide juncture between descriptive narratives and empirical analyses.

Acknowledgment

This article is based on research supported by the Social Sciences and Humanities Research Council of Canada.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bloom, B.S., & Broder, L.J. (1950). *Problem-solving processes of college students.* Chicago: University of Chicago Press.

Branca, N.A. (1980). Problem solving as a goal, process, and basic skill. In S. Krulik & R.E. Reys (Eds.), *Problem solving in school mathematics* (pp. 2-8). Reston, VA: National Council of Teachers of Mathematics.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Lindquist, M.M., & Reys, R.E. (1981). National assessment. In E. Fennema (Ed.), *Mathematics education research: Implications for the 80's* (pp. 22-38). Reston, VA: National Council of Teachers of Mathematics.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(2), 2-10.

Cordua, G.D., McGraw, K.O., & Drabman, R.S. (1979). Doctor or nurse: Children's perceptions of sex typed occupations. *Child Development, 50*, 590-593.

Davis, G.A. (1973). *Psychology of problem solving: Theory and practice.* New York: Basic Books.

Ericcson, C.W., & Staszewski, J.J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahar & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, NJ: Erlbaum.

Feigenbaum, E.A. (1989). What hath Simon wrought? In D. Klahar & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 163-182). Hillsdale, NJ: Erlbaum.

Gagne, R.M. (1983). Some issues in the psychology of mathematics instruction. *Journal for Research in Mathematics Education, 14*, 7-18.

Glaser, R. (1989). Expertise and learning: How do we think about instructional processes now that we have discovered knowledge structures? In D. Klahar & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 269-282). Hillsdale, NJ: Erlbaum.

Greeno, J.G. (1980). Trends in the theory of knowledge for problem solving. In D.T. Tuma & R. Rief (Eds.), *Problem solving and education: Issues in teaching and research* (pp. 9- 23). Hillsdale, NJ: Erlbaum.

Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episodic structure of algebra story problems. *Cognition and Instruction, 6*, 223-283.

Hyde, S.H., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Kimball, M.M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198-214.

Kintsch. W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109-129.

Larkin, J.G., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Models of competence involving physics problems. *Cognitive Science, 4*, 317-345.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* New York: Cambridge University Press.

Lester, F.K. (1980). Research on mathematical problem solving. In R.J. Shumway (Ed.), *Research in mathematics education* (pp. 286-323). Reston, VA: National Council of Teachers of Mathematics.

Lester, F.K., Garofalo, J., & Kroll, D.L. (1989). Self-confidence, interests, beliefs, and metacognition: Key influences on problem-solving behavior. In D.B. McLeod & V.M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 75-85). New York: Springer-Verlag.

Lewis, A.B., & Mayer, R.E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*, 363-371.

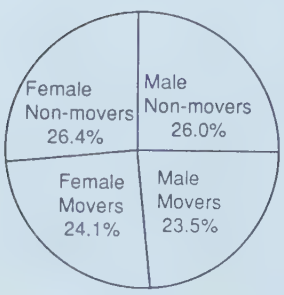
Mayer, R.E. (1985). Mathematical ability. In R.J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 127-150). New York: Freeman.

Mayer, R.E. (1986). Mathematics. In R.F. Dillon & R.J. Sternberg (Eds.), *Cognition and instruction* (pp. 127-154). New York: Academic Press.
Mayer, R.E. Larkin, J.H., & Kadane, J.B. (1984). A cognitive analysis of mathematical problem-solving ability. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (vol. 2, pp. 231-273). Hillsdale, NJ: Erlbaum.
McLeod, D.B. (1989). The role of affect in mathematical problem solving. In D.B. McLeod & V.M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 20-36). New York: Springer-Verlag.
Paivio, A. (1978). On exploring visual knowledge. In B.S. Randhawa & W.E. Coffman (Eds.), *Visual learning, thinking, and communication* (pp. 113-131). New York: Academic Press.
Polya, G. (1957). *How to solve it: A new aspect of mathematical method.* Garden City. NY: Doubleday Anchor Books.
Polya, G. (1962). *Mathematical discovery* (vol. 1). New York: Wiley.
Randhawa, B.S., & Beamer, J. E. (June, 1990). *Learning and instructional processes in grade four classrooms.* Paper presented at the Canadian Society for the Study of Education, Victoria.
Randhawa, B.S., Ng, C.F., & Beamer, J.E. (1990, June). *Mathematical competence: Personal and social influences.* Paper presented at the annual convention of the Canadian Psychological Association, Ottawa.
Russell, R.L., & Ginsberg, H.P. (1984). Cognitive analysis of children's mathematical difficulties. *Cognition and Instruction, 1,* 217-244.
Scandura, J.M., Durnin, J.H., & Wulfeck II, W.H. (1977). Higher-order rule categorization of heuristics for compass and straight-edge constructions in geometry. In J.M. Scandura (Ed.), *Problem solving: A structural/process approach with instructional implications* (pp. 165-199). New York: Academic Press.
Shepard, R.N. (1978). Externalization of mental images and the act of creation. In B.S. Randhawa & W.E. Coffman (Eds.), *Visual learning, thinking, and communication* (pp. 133-189). New York: Academic Press.
Signorella, M.L., & Liben, L.S. (1984). Recall and reconstruction of gender-related pictures: Effects of attitude, task difficulty, and age. *Child Development, 55,* 393-405.
Snyder, M. (1987). *Public appearances/private realities: The psychology of self-monitoring.* New York: Freeman.
Sternberg, R.J., & Salter, W. (1982). Conceptions of intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 3-28). Cambridge: Cambridge University Press.
Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112,* 639-661.
Whitehead, A.N. (1929). *The aims of education.* New York: MacMillan.
Wong, N.Y. (1988). Effects of self-monitoring and reinforcement on problem solving performance. *Educational Psychology, 8,* 153-159.
Zigler, E., & Seitz, V. (1982). Social policy and intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 586-641). Cambridge: Cambridge University Press.

Appendix A
Core Problems and the Summary of Scoring Scheme for Each Problem

Problems

1. Population 5 years of age and over for Canada, 1981
 Total population 5 years of age and over = 22,280,070

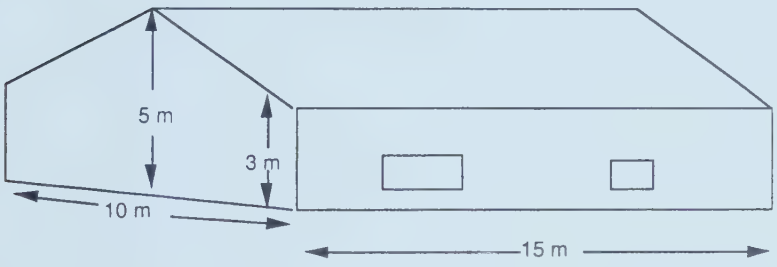


Source: Census statistics, 1981.

Movers are defined as persons who, on Census Day, were living in a different dwelling than the one occupied 5 years earlier.

(a) Calculate the number of female movers.
(b) What proportion of males were movers?
(c) How many more female movers than male movers were there?

2. An identical item is on sale at two stores. At store A, the sale price is 24% off the regular price. At Store B, the sale price is 10% off the price which has already been marked down by 15%. Assuming that the regular prices at both stores are the same, which store has a better offer?
3. A car rental company offers two rental rates for a medium-size compact car. Either $35 daily with unlimited mileage, or $20 daily plus $0.15 per km. Which rate is better to rent a car for a day? Why?
4. Susan paid $600 US for a stereo set when she visited her uncle in San Francisco last week. If the exchange rate was $1 Canadian to $0.85 US, how much did her stereo cost in Canadian dollars?



5. You are going to paint the house shown below with one coat of paint. It has a 1m by 2m picture window at the front and six smaller windows measuring 1m by 1m on the sides and the back. The roof will not be painted.
 A rule of thumb that painters use is one litre of paint for $9.3m^2$ of surface.
 The regular price for a 1-litre can is $10.95 but now on sale at 25% discount.
 The regular price for a 4-litre can is $27.95.
 If you cannot use any unused portion of the paint in a can for anything else, what will be the lowest cost to paint the house?

Details of individual score allocations are available from the authors.

Scoring Scheme[a]

Problem	*Process*	*Computations*	*Total*
1.	7	2	9
2.	6	2	8
3.	4	1	5
4.	3	1	4
5.	14	3	17

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 233-245*

David Bateson
University of British Columbia

Psychometric and Philosophic Problems in "Authentic" Assessment: Performance Tasks and Portfolios

The concept of "authentic" measurement/assessment/evaluation, which has developed and grown from gross misuse and a perception of unfairness and a lack of validity in past measurement, assessment, and evaluation practices, has been rapidly and enthusiastically accepted by many educators. Use of "portfolios" and "performance" tasks has exponentially increased as educators seek to be as "fair" as possible, and to link measurement, assessment, and evaluation more closely to cognition, curriculum, and instruction. However, these "newer" methodologies are fraught with all the problems and pitfalls that have existed throughtout measurement, assessment, and evaluation history. The rigorous employment of high standards of reliability and validity, albeit in possibly different forms from past practice, is equally essential in the use of "authentic" methods and practice as it was in past practice. Much of what has been learned over the long history of non-"authentic" measurement can provide at least the basis for possible solutions to present problems.

Le concepte de mesure/contrôle/évaluation "authentique" qui a pris forme et qui s'est propagé à partir d'un flagrant abus d'emploi, d'une perception d'injustice, et d'un manque de validité dans l'usage antérieur des moyens de mesure, de contrôle, et d'évaluation a été accepté rapidement et avec enthousiasme par beaucoup d'éducateurs et d'éducatrices. Puisque les enseignant(e)s cherchent à être aussi "justes" que possible vis-à-vis l'évaluation, le testing, et la mesure, et de les lier plus intimement à la cognition, au curriculum, et à l'instruction, l'utilisation de "portfolios" et la détermination des habiletés de "performance" s'est accrue de façon exponentielle dans le niveau scolaire. Cependant, ces méthodologies dites "nouvelles" sont chargées de toutes sortes de problèmes et de pièges qui ont toujours existé dans l'histoire de la mesure, du contrôle, et de l'évaluation. L'emploi rigoureux de standards élevés en justesse et en fiabilité, bien qu'il soit utilisé de façons différentes que par le passé, est également essentiel au développement et à l'utilisation des méthodes et des pratiques d'évaluation "authentiques." Beaucoup de ce qui a été appris dans la longue histoire de la mesure non-"authentique" pourrait servir de base à des solutions possibles aux problèmes d'aujourd'hui dans les domaines de l'évaluation, du contrôle, et de la mesure.

During the last three decades, educators (particularly classroom teachers and school administrators) became more and more disillusioned with what they perceived to be inappropriate use of standardized, right-or-wrong, pencil-and-paper, multiple-choice tests. Teachers, administrators, and politicians, in a desire to be as objective (read "fair") as possible, but without adequate knowledge or training about measurement, often grossly misused information from individual tests and testing programs. Although they pointed out that selection-type items, particularly multiple-choice, could not measure many of the

David Bateson is an associate professor in the Department of Mathematics and Science Education and a research associate with both the Educational Measurement Research Group and the Centre for Applied Studies in Evaluation. He specializes in science education, measurement, and evaluation.

important goals of education such as the ability to write (as opposed to edit) and the ability to employ complex thinking skills, they continued to depend on such items for lack of any acceptable competitor. The growing negative feelings built up over the years, until the end of the 1980s, when the education community was desperate for change, and a revolution against standardized, selection-type tests was inevitable.

Into the situation came the champions of "authentic" assessment. By its very name, authentic assessment donned a mantle of protective mail. More recently the concepts of "performance tasks" and "portfolios" have also been gathered under authentic's protective wing.

Authentic Assessment

The concept of authentic measurement was first proposed by Archbald and Newmann (1988) and more recently expanded in Berlack et al. (1992). It now dominates discussion of student assessment. A complete shift in basic beliefs and values in measurement has occurred (Kleinsasser, Horsch, & Tastad, 1993) that might be likened to a paradigm shift on the part of many educators, administrators, and politicians.

As an indicator of the extent and effect of this movement, an examination of the program for the annual meeting of the National Council on Measurement in Education held in Washington, DC in 1987, the year before the Archbald and Newmmann document was produced, showed that not one of the 53 session titles contained the words *authentic, performance,* or *portfolio* as they have come to be known in terms of assessment. In 1990, when many psychometricians may still have been hoping that the authentic movement was merely a flash in the pan, there were three papers: two about performance and one about portfolios, but all three dealt with evaluation of teachers rather than assessment of students. By 1993 three of the six presession training courses, and 15 of the regular sessions concentrated on aspects of the authentic movement. A quick glance at the American Educational Research Association programs shows a similar trend.

Change seen in the contents of scholarly discourse reflects the changes in practice in schools. In more and more jurisdictions test scores and grades that in the past parents have thought they understood, are being replaced by anecdotal reports and "narrative profiles" (Moss et al., 1992). The thrust in school reporting has gone from norm-referenced information about the relative ranking of students in various subject areas, to criterion-referenced information about what a given student "can do." Often such "can do" information is given without reference to what a child at that age and/or grade should be expected to do. Faced with this type of reporting, parents complain that they do not understand such information, and they desperately want to know "Where does my child stand?" However, so complete is the transfer to criterion-referenced and can-do reporting, that some schools and teachers are refusing steadfastly to respond to any norm-referenced concerns of the community they serve.

Despite its widespread use and acceptance, it has been difficult to decipher what the term *authentic assessment* means, particularly in relation to current practices. In discussions with university and teacher training personnel, Ministry of Education officials, administrators, and many classroom teachers, it

becomes obvious that there are multiple meanings for the term, depending on the individual questioned and the particular practices employed. However, a common element in most definitions seems to be *any measurement technique that is not a strong component of mainstream, traditional testing and assessment practices up to the date of the coining of the term*. Obviously such a definition or understanding of the term is neither appropriate nor useful in trying to come to grips with what authentic means when used in conjunction with assessment, evaluation, or measurement. From Archbald and Newmann's (1988) original work, authenticity should have two key features:

1. reporting should present the level of mastery of tasks and the nature of the task used to measure that mastery [criterion referencing], and
2. tasks presented to students must be worthwhile, significant and meaningful. (p. 1)

The document goes on to present three principal characteristics of the tasks:

1. production of discourse, things, performances;
2. flexible use of time; and
3. collaboration (pp. 3-4),

and subsequent documents have modified the principal characteristics:

1. collaborations,
2. access to tools and resources,
3. worker discretion or opportunity for ownership, and
4. flexible use of time. (Newmann & Archbald, 1992, pp. 78-79)

Shepherd (1991) makes the best distinction between typical traditional test items and authentic test items: "The [authentic] assessment tasks themselves are real instances of estimators of actual learning goals" (p. 21). Typical traditional test items require the test interpreter to infer from the item or a group of items competence in a specified domain of achievement or interest, whereas authentic tasks are the domain of achievement or interest; interpretation of the item or instrument is not necessary because the assessment is the product itself.

Defining the assessment this way neatly protects it from any questioning about important issues such as are contained in the traditional psychometric views of reliability and validity (e.g., Bateson, 1992). Conversely, any method that is not deemed authentic is, by logic, not authentic, not reliable, not trustworthy, and/or not genuine. One cannot imagine a classroom teacher telling a parent, "Well, we want to do some nonauthentic assessment to alert us to possible problems in reading."

A problem of definition also seems to occur for the related term *performance task*. Unfortunately, it appears that the term *performance* has come to mean, to a large extent, "writing an essay about ..." The National Council on Measurement in Education Presession Course H (Linn, 1991) conducted by the CRESST group, demonstrated scoring rubrics for essays in English, social sciences, mathematics, biology, and physical sciences.

It is difficult, though not impossible, to construe an essay in or about mathematics as being a real or authentic instance of student achievement in mathematics! However, such a trap is relatively easy to fall into because

measurement issues surrounding the scoring of essays have been the subject of development and investigation for many years. The techniques of developing concise, well-defined scales based on specific criteria that are illustrated with exemplars or "anchor papers" representing the range of responses in the total set of responses, and carefully training and retraining of scorers/markers using common illustrative papers, have led to generally acceptable levels of marker agreement in many applications. In the absence of other proven alternative techniques, essay writing has become the primary manifestation of performance assessment.

Despite the comfort of psychometricians with carefully constructed procedures for scoring essays, and the seeming elegance of analyzing and reporting essay results, measurement scholars must not try to misrepresent what is being accomplished through the application of essay writing in the guise of performance tasks: writing an essay is not congruent with doing science just as writing an essay about skiing provides no information about actual skiing abilities. In addition, psychometricians should not become too satisfied with using written expression samples to evaluate students. Despite the increased use of these type of instruments and measures over many years, many of the basic reliability and validity problems have still not been solved. "The development of performance assessments in the early years of the 1980s contained a major set of hurdles that remain to be crossed by those who choose to enter the assessment field" (Chapman & Kerins, 1993, p. 2).

Portfolio Assessment

Theoretical principles have been presented that treat portfolios similarly to writing samples (Reckase, 1993). Four- or five-point scales used to rate the portfolio on various evaluative components are developed, specific written criteria for each scale point are defined and agreed on, and exemplars of each scale point are collected. Raters of portfolios are trained through extensive examination of the scale and each point on the scale, group scoring of common portfolios accompanied by extensive discussion about the judgments made, and teams of scorers working together and sharing ideas.

Despite the theoretical appeal of such procedures and the fact that their application has resulted in high interrater reliability in previous assessments of written expression, studies of large-scale implementations of portfolio assessments have shown that low interrater reliability indices result (Koretz, Stecher, & Deibert, 1993). In the study of the Vermont implementation, the 38 interrater reliability indices for judgments about various aspects and subjects of the portfolios ranged from a low of .28 for the grade 4 "language of math" composite score to a high of .57 for the grade 8 writing "usage" score, and only three of the indices reached .50. The mean of the 38 indices was .37 with a standard deviation of .09 (Koretz, McCaffrey, Klien, Bell, & Stecher, 1993).

One of the major differences between past written expression assessments and modern portfolio assessments is that although the stimulus and response form of the written expression studies has usually been identical for all students, the content of portfolios of work by various students may differ in the extreme. In fact administrators and teachers have been told that "portfolio assessment can take a variety of shapes in a variety of circumstances ... which will permit a range of options, rather than a singular definition" (Fisher, 1991,

p. 10). Although one student's portfolio may contain six examples of writing, a single portfolio may contain a set of pictures of a science fair display, an audiotape of the student reading a narrative passage, and four ribbons from a district track meet. Traditional assessment and testing methods cannot help in developing and validating methods of comparing such diverse data because traditional comparative methods have been based on standardization. Without some standardization evaluating portfolios is as difficult as its parallel problem, the proverbial comparison of apples and oranges.

Although many problems remain to be addressed regarding the implementation of authentic measurement tasks, at least one appears to have been solved, at least for the moment. The standard in terms of methods for developing reliability indices of performance tasks seems to have settled on studies employing generalizability coefficients (Bateson & Blackmore, 1983; Linn, 1991; Trevisan, 1991; Welch, 1993) that were first proposed by Cronbach, Gleser, Nanda, and Rajaratnam (1972), and later developed by Brennan (1983) and Crick and Brennan (1983). Analysis of variance techniques are used to partial the different sources of variance (students, tasks, occasion, scorer/marker, etc., and all the accompanying interactions) and manipulation of the "explained" variables can be used to minimize error.

A Case Example: The BC Science Assessment

The province of British Columbia has joined the states of Connecticut, Vermont, California, and Maryland in radically changing accountability and evaluation systems that have traditionally relied on standardized multiple-choice testing and turning instead to systems that rely more on performance-based tasks and criterion-referenced interpretations (Frechtling, 1991).

With the implementation of the primary program under the auspices of the Year 2000 program (British Columbia Ministry of Education, 1991), the BC Ministry has directed that reporting to parents be essentially interview-type anecdotal records, and have further directed that the use of checklists and standardized norm-referenced tests is inappropriate for primary students. In addition, although the Ministry initially proposed "signposts" and currently has contracts with researchers to develop "reference sets," Ministry expectations are interpreted by primary teachers to mean that even comments based on any criterion-referenced standard are inappropriate; teachers should report only what a student can do without reference to what is normally expected for pupils at that age and grade (British Columbia Ministry of Education, 1991).

The BC Science Assessment, which won the 1993 Division H American Educational Research Association publication award for the best overall instructional program evaluation, was designed specifically to show the benefits of blending traditional and newer forms of assessment practice (Bateson, Erickson, Gaskell, & Wideen, 1992). The overall evaluation used methods and techniques which ranged from strict application of traditional quantitative methods (Bateson, Anderson, et al., 1992); through analyses of students' writing in response to both print and video stimuli regarding socioscientific issues (Gaskell, Fleming, Fountain, & Ojelel, 1992); through description and analyses of closed and free responses to simple science-oriented tasks using materials, and the coding/evaluation of more complex and extensive hands-on investigations (Erickson et al., 1992); to employing videotape, structured observa-

tions, and interviews of administrators, science teachers, and students doing science activities to present the context of science teaching and learning (Wideen et al., 1992).

Although the Ministry of Education usually receives a significant number of complaints from teachers and administrators about the perceived validity and value of each assessment as it is being conducted, for the 1991 Science Assessment, which demonstrated an obvious merging of the many data collection and interpretation methods and techniques discussed above, not only did Ministry officials not receive one single complaint from the field, but they received many laudatory letters. Such acceptance by teachers of external testing was without precedent in British Columbia. For the 1991 BC Science Assessment, teachers perceived the methods and techniques to be appropriate and valid, and many described them as "authentic."

It is essential that more highly visible projects of this nature be conducted. Teachers must see and become part of projects that use a mutually supportive combination of the older, tried-and-true methods and techniques, and the more recently developed methods with which researchers are less secure. It is a time when large-scale measurement, assessment, and evaluation specialists must take some risks. It must be demonstrated, as it was in the 1991 BC Science Assessment, that the best decisions will be made if the data on which the decisions are based come from multiple, complementary methods using a variety of data types. It must be clearly demonstrated that if one relies on a single method or a single source of data, there is a much greater potential for erroneous and invalid decisions. Such demonstrations can have a high impact on teachers and will do much to influence practice, particularly when large numbers of them can be directly involved in the process.

Current Status of Knowledge of Assessment Practice

Many well-thought-out and well-researched techniques for developing "authentic" measures have been published (Shavelson & Baxter, 1992; Swain, 1989). However, these articles/papers have been presented through meetings and journals that are not popular reading for classroom teachers or even administrators. It is essential that messages such as those described above somehow get conveyed to the members of the teaching profession who might be applying such techniques in their classrooms.

The accepted techniques and standards for testing using traditional methods, which have substantially been of a quantitative nature, are well documented in a great number of presentations, articles, and texts. These methods form the foundation not only of the teacher training courses that most teachers have taken in the past, but also of the testing traditions and "folk knowledge" of schools. However, more recently attention has also been directed at qualitative methods of gathering data and research. Despite having entered the mainstream of accepted research methods long after the acceptance of quantitative methods, there has been a rapid acceptance of qualitative research, particularly over the last decade. A major factor in such "softer" methods being accepted has to do with the various challenges to the integrity of the qualitative methodologies that have occurred.

In response to such challenges, researchers promoting the techniques of interviews, observations, and participant observations have developed per-

suasive, logical arguments for their particular procedures and have presented rules of evidence that can be employed to evaluate the procedures. As such, the validity of many models of qualitative research has been established, and these proven qualitative methods and techniques are well suited to use by classroom teachers to provide better data on which to make their educational decisions about students.

The standards for acceptability of both quantitative and qualitative methods, however, are not widely known and understood by most educators, particularly classroom teachers, who are not aware of the many methods and techniques that are available to them. University preservice training has generally failed dismally in preparing teachers for testing, assessment, and evaluation tasks which they must undertake in their classrooms. For the minority of students who actually take a course in testing, measurement, and/or evaluation, the courses almost exclusively deal with the development, evaluation, and use of paper-and-pencil tests to the exclusion of most other measurement and evaluation topics. Even many of the available texts only pay lip service to the topics of alternative assessment procedures.

Without appropriate knowledge and training, teachers have fallen back on the folk knowledge of testing: What was done to them (usually inappropriately). Even the language of psychometrics is usually misunderstood; to most teachers, "validity" is a personal synonym for "good" or "something I like," as opposed to the meanings that test constructors and psychometricians typically ascribe to validity. Most teachers would be amazed if they could understand Messick's (1989) detailed and complex argument concerning validity whose basic message of consequences makes so much common sense. In trying to solve pressing, day-to-day problems of assessment and measurement in their classrooms, teachers and administrators have tried to mimic the techniques and instruments used by large-scale testing and assessment programs that have been the darling of most psychometricians. What they fail to realize is that methods which are generally appropriate for making some types of decisions about groups and systems, or that are designed merely to "red-flag" individuals, are usually not appropriate for making high-stakes classroom decisions about individuals. Teachers and administrators for too long have been the victims and have seen the students for whom they care become victims of inappropriate measurement practices.

In reaction, they have wholeheartedly joined the authentic revolution. Although the established processes and methods of standardized, paper-and-pencil, black and white, selection-type testing have been researched and developed well and have served the purposes of making policy decisions about groups at relatively low cost in the past, the present mushrooming of interest in authentic, performance, and portfolio procedures dictates that the focus of resources, training, and research in the testing/measurement/evaluation community must change.

Maguire (1992) makes the case that authentic assessment must be firmly grounded in educational philosophy and, by implication, cognitive psychology and uses the example of the SOLO taxonomy (Biggs & Collis, 1982) as firmly grounded assessment methodology. Similar calls for firm grounding not only in educational philosophy and cognitive psychology, but also in knowledge of

the structure of the subject matter have been made by other authors (Linn, Baker, & Dunbar, 1991; Moss, 1992). In describing validation criteria, Moss (1992) refers to:

> Content quality: the extent to which the content is consistent with the best current understanding of the field as evidenced by systematic judgments of the quality of the tasks from subject matter specialists and evidence about the ways in which students interpret the content. (p. 248)

Unfortunately, few examples of such sophisticated analysis can be found in classroom, school, district, or system practice. If authentic measures that are being used could point to and provide evidence for such grounding, the issues of validity could more easily be addressed. However, it is usually the case that the derivation or development of such measures is unknown, and the researcher interested in the validity of the measures is left to attempt to infer the underlying educational philosophy and the cognitive psychology bases.

Performance assessment in science is one area where the tasks that have been developed and used to date are well grounded in at least the philosophy of science, and they generally allow students opportunities to demonstrate a range of cognitive complexity. However, few tasks are of quality, and one reason is the extreme difficulty of generating and refining the tasks and the accompanying data collection and scoring/coding criteria. As one group of researchers has described the situation:

> We have found the process of creating performance assessments ... to be time consuming, requiring considerable scientific and technological know-how. Development of quality performance assessments requires multiple iterations through a sequence of development, tryouts with students (getting their thoughts and comments), and revision. Short-circuiting this process leads to ill-conceived and poorly constructed assessments. (Shavelson & Baxter, 1992, p. 231)

The fact that good performance tasks are extremely difficult to develop means that few good performance tasks are available. To compound the problem, because most performance tasks are purportedly designed to assess problem solving and higher level thinking skills, it is essential that the task used for any assessment be novel to the students being assessed. If the students have seen and done the task previously, their performance on the assessment merely becomes a regurgitation of what they remember from the last time they did it. The fact that performance/items must be piloted, often on several occasions, means that the security of the item that is required in order to ensure that the item is measuring what was, or is intended, rather than just recall or replication of previously taught solutions, is compromised.

Teachers are excellent judges of quality performance tasks. As previously identified in the discussion of the definition of authentic measures, because the tasks are designed to be the domain of interest, the tasks are quickly recognized and adopted by teachers as good teaching and learning activities. With the amazing "moccasin telegraph" that exists among teachers, it requires little time for most teachers to become aware of such good tasks. If the item escapes wide distribution during piloting, as soon as it is used to any extent on any

assessment it usually becomes widely known and used by teachers in the course of instruction, and so becomes worthless as an assessment device.

A prime example of such a task is the "paper towels" task that probably originated with the Assessment Performance Unit in Britain, but has now been used in at least 11 jurisdictions on at least four continents and has been the subject of many scholarly presentations, articles, and reports. Teachers who have been involved with any administration, review, scoring, or interpretation tasks associated with the assessments have recognized the elegance of the paper towel task and have used it for teaching purposes as opposed to assessment purposes. Because many students have now interacted with this task, this formerly ideal item/task is now totally useless for assessment of anything but recall of information.

Linn et al. (1991) have proposed eight criteria for judging performance assessments:

1. Intended and unintended consequences,
2. Transfer and generalizability,
3. Fairness,
4. Cognitive complexity of student processes,
5. Meaningfulness of the problems,
6. Content quality,
7. Content comprehensiveness of coverage, and
8. Cost and efficiency. (p. 20)

Transfer and Generalizability Issues

It is difficult to argue against any of the criteria proposed, but merely identifying the general criteria does not define exactly what each criterion means or the evidence that would be acceptable to come to a judgment about it. Without discussing the relative merits of possible ways of implementing the proposed set of criteria as a whole, it is useful to focus on the second criterion, transfer and generalizability. In an expansion of the criteria, Linn et al. (1991) state that the second criterion should be applied as "the degree to which performance on specific tasks transfers" (p. 20). Moss (1992) further expands the meaning of this criterion to include "the extent to which results can be generalized across raters and tasks (reliability, in the traditional sense) and from the specific assessment task to the broader domain of achievement" (p. 248).

With this issue of transfer and generalizability in mind, attempts have been made to standardize performance tasks and to develop generic tasks that purport to measure a generic skill or process without regard to the knowledge and/or interest of the subject matter context (e.g., Jorgensen, 1993). However, such attempts have to date been singularly unsuccessful. Numerous studies of assessments which have attempted to use both subject-independent and structured performance tasks have shown that the context of the performance task is all-important; that generalizable or generic skills and processes that students may possess independent of content, while theoretically appealing, are not possible to operationalize in an assessment situation (Bunch & Littlefair, 1988; Erickson, et al., 1992; Haladyna & Sloat, 1993; Shavelson & Baxter, 1992; Shavelson, Baxter, Pine, & Yura, 1990; Welch, 1993). Even with the scoring of essays, which are the performance tasks which present the fewest problems for psychometricians, "topics are not interchangeable and some students write

better on one prompt while other students write better on other prompts" (Welch, 1993, p. 11). Further, it is also certain that although the writing prompt is identical for all examinees, each examinee's understanding of, and knowledge and interest in, any given task will be somewhat different; "the idea that such an assessment is a fair measure of each respondent's performance is thus open to challenge" (Bunch & Littlefair, 1988, p. 175).

Although it has been possible to develop selection-type testing items, such as multiple-choice, that have high inter-item correlations, and therefore can be inferred to all measure approximately the same construct, similar inter-item correlations have not been found for performance tasks. Tasks which appear to be similar, or appear to require the identical skill applied in the same way but in a different context have shown very low, and often nonsignificant inter-item correlations.

In addition, the nature of performance tasks, particularly those designed to measure the use of higher level thinking and complex problem solving skills, dictates that many of the variables that are fixed in traditional standardized tests are purposely let free to vary in performance tasks. Not the least of the free variables is often the "correct" answer. Whereas in traditional standardized tests there is almost always one single answer that is scored as correct, in performance tasks there may be many more or less appropriate and successful ways to approach the problem. Such being the case, the use of internal consistency measures of reliability will necessarily result in coefficients that are much lower than those obtained from typical standardized multiple-choice tests. For performance tasks, this standard must change—given the nature of performance tasks, internal consistency cannot be expected to be strong and one could argue that internal consistency measures are not even appropriate for use in performance assessments. "Although efficiency, reliability, and comparability are important issues that cannot be ignored in new forms of assessment, they should not be the only, or even the primary criteria to judge the quality and usefulness of an assessment" (Linn et al., 1991, p. 16).

Psychometricians must broaden their typical view of reliability measured through internal consistency to include dependability measures based on inter- (or intra-) rater/observer/scorer analyses. There are also persuasive arguments that even if internal consistency indices are used as the standard of reliability, low indices are worth the price of the information gained.

> The beauty and potential use of the performance assessment is to obtain outcome data on valued achievement targets difficult to assess with standardized achievement tests. The potential gain in further achievement data is obtained at a loss in the magnitude of the reliability coefficient. (Trevisan, 1991, p. 3)

Because most performance tasks, by definition of being authentic measures, actually are the domain of interest, such tasks have great "face validity." Teachers speak highly of them and often will compete to attempt to get their class involved in the assessment. Such competition to be included is diametrically opposite to teacher behaviors in traditional assessment. The perception of such tasks is that they truly measure what should be valued in education. With such wide acceptance by teachers who are traditionally very critical, it is difficult to believe that these performance tasks should be judged to be less

than desirable simply because they are not able to be transferred and do not represent generalizable skills or processes. Because no tasks that have been developed to date can be acceptable based on the proposed criteria of being transferable or generalizable, it may be necessary to reconsider and reconceptualize how performance tasks should be judged.

Future Assessments

Some visions of the future in testing have included alarming scenarios: "NRTs [norm-referenced tests] and CRTs [criterion-referenced tests] have become extinct—as have the spotted owl, African elephant, great white shark, bottle nosed dolphin, and numerous species of birds and insects" (Hansen, 1993, p. 3). If such is to be the case and useful psychometric techniques that to a large extent have been well understood are fated for oblivion, only radical action on the part of psychometricians to demonstrate the benefits of such older methods, similar to radical actions taken by environmentalists, may be necessary. Stiggens (1991) argues that the changes occurring are noteworthy: "Our current assessment upheaval is not simply the latest fad to sweep the education scene. Rather, it signals the end of a 60-year era of educational assessment and our passage into a whole new era" (p. 263).

However, others consider that the present thinking about performance testing and portfolios is just another wave that if merely ignored will recede to the background (Wolmut, 1993).

This type of ostrich-like thinking and action on the part of measurement and evaluation specialists has led to the unfortunate wholesale acceptance of techniques dubbed authentic in today's classrooms. Many psychometricians and others in the measurement community have not responded well to past criticisms, and many are now treating the authentic movement with disdain. They have failed to provide widespread, useful guidance for classroom practitioners in the past, and have more recently failed to respond appropriately to the rapid acceptance by classroom teachers of an enticing solution to what these teachers have seen as the testing monster.

Rather than joining the rush in order to ensure that the development of the authentic movement proceeds in an orderly fashion, many knowledgeable measurement scholars have either ignored what is happening at the classroom level, or have steadfastly but vainly decried what they see as a threat to their comfortable position. They have vigorously attacked the whole concept of authenticity. Although they may have excellent grounds for their attacks, school practitioners are not listening. Fortunately, the tide is turning, and more respected measurement individuals have turned their attention in constructive rather than destructive ways to the changes caused by the ongoing authentic revolution. They are "getting on the bandwagon" rather than trying to "derail the train," and their actions will most assuredly assist in better practices being instituted.

References

Archbald, D.A., & Newmann, F.M. (1988). *Beyond standardized testing: Assessing authentic achievement in the secondary school*. Reston, VA: National Association of Secondary Principals.

Bateson, D.J. (1992). Adding authenticity to sound measurement practice. *Research Forum, 1*(9), 4-8.

Bateson, D.J., Anderson, J., Brigden, S., Day, E., Deeter, B., Eberle, C., Gurney, B., & McConnell, V. (1992). *British Columbia assessment of science 1991 technical report I: Classical component*. Victoria, BC: Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights.

Bateson, D., & Blackmore, D. (1983, June). *Application of generalizability theory to the number of markers question in scoring of written expression*. Paper presented at the annual meeting of the Canadian Society for Studies in Education, Vancouver, BC.

Bateson, D., Erickson, G., Gaskell, P.J., & Wideen, M. (1992). *British Columbia assessment of science (1991) provincial report*. Victoria, BC: Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights.

Berlack, H., Newmann, F. M., Adams, E., Archbald, D.A., Burgess, T., Raven, J., & Romberg, T.A. (1992). *Toward a new science of educational testing and assessment*. Albany, NY: State University of New York Press.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Brennan, R.L. (1983). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.

British Columbia Ministry of Education. (1989). *Year 2000: A curriculum and assessment framework for the future*. Victoria, BC: Author.

British Columbia Ministry of Education. (1991). *Primary program*. Victoria, BC: Author.

Bunch, M.B., & Littlefair, W. (1988). *Total score reliability in large-scale writing assessment*. Paper presented at the Education Commission of the States, Boulder.

Chapman, C.W., & Kerins, C.T. (1993, April). *How do we report useful information to schools based on authentic assessment*. Paper presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Atlanta.

Crick, J.E., & Brennan, R.L. (1983). *Manual for GENOVA: A generalized analysis of variance system*. Albany, NY: American College Testing.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurement*. New York: Wiley.

Erickson, G., Bartley, A.W., Blake, L., Carlisle, R.W., Meyer, K., & Stavey, R. (1992). *British Columbia assessment of science (1991) technical report II: Student performance component*. Victoria, BC: Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights.

Fisher, D. (1991). Portfolio assessment: A framework for a range of options. *Research Forum, 1*(9), 9-12.

Frechtling, J.A. (1991). Performance assessment: Moonstruck or the real thing? *Educational Measurement: Issues and Practice, 10*(4), 23-25.

Gaskell, P.J., Fleming, R., Fountain, R., & Ojelel, A. (1992). *British Columbia assessment of science (1991) Technical report III. Socioscientific issues component*. Victoria, BC: Ministry of Education.

Hansen, J.B. (1993, April). *Assessment in the year 2001: The darkness and the light*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta.

Haladyna, T.M., & Sloat, E. (1993, April). *The influence of context in reporting achievement test scores*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta.

Jorgensen, M. (1993, April). *Authentic assessment for multiple users*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta.

Kleinsasser, A., Horsch, E., & Tastad, S. (1993, April). *Walking the talk: Moving from a testing culture to an assessment culture*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont Portfolio Assessment Program: Interim report*. RAND Institute.

Koretz, D., Stecher, B., & Deibert, E. (1993, April). *The Vermont portfolio assessment program: Interim report on implementation and impact, 1991-92 school year*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta.

Linn, R.L. (1991, April). *Issues in large scale performance assessment*. Course H training session held at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, Illinois.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Maguire, T. (1992). Grounded authentic assessment and teacher education. In D.Bateson (Ed.) *Classroom testing in Canada: Proceedings of the second Canadian conference on classroom testing*. Vancouver: Centre for Applied Studies in Evaluation.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103) New York: Macmillan.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229-258.

Moss, P.A., Beck, J.S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R. (1992). Portfolios, accountability, and the interpretive approach to validity. *Educational Measurement: Issues and Practice, 11*(3), 12-21.

Newmann, F.M., & Archbald, D.A. (1992). The nature of authentic academic achievement. In A.R. Tom (Ed.), *Toward a new science of educational testing and assessment*. Albany, NY: State University of New York Press.

Reckase, M.D. (1993, April). *Portfolio assessment: A theoretical prediction of measurement properties.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Shavelson, R.J., & Baxter, G.P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49*(8), 20-25.

Shavelson, R.J., Baxter, G.P., & Pine, J., & Yura, J. (1990, April). *New technologies for large-scale science assessments: Instruments of educational reform.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Shepard, L. (1991). Interview on assessment issues with Lorie Shepard. *Educational Reseearcher, 20*(2), 21-23.

Stiggens, R.J. (1991). Facing the challenges of a new era of educational assessment. *Applied Measurement in Education, 4(4), 263-273.*

Swain, J.R.L. (1989). The development of a framework for the assessment of process skills in a graded assessment in science project. *International Journal of Science Education, 11*(2), 26-38.

Trevisan, M.S. (1991, April). *Reliability of performance assessments: Let us make sure we account for the errors.* Paper presented at the annual meeting of the National Association of Test Directors, Chicago.

Welch, C.J. (1993, April). *Issues in developing and scoring performance assessments.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Wideen, M., Mackinnon, A., O'Shea, T., Wild, R., Shapson, S., Day, E., Pye, I., Moon, B., Cusack, S., Chin, P., & Pye, K. (1992). *British Columbia assessment of science (1991) technical report IV: Context for science component.* Victoria, BC: Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights.

Wolmut, P. (1993, April). *Do we know where we're going? Have we been there before? Riding the measurement waves.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta.

The Alberta Journal of Educational Research *Vol. XL, No. 2, June 1994, 247-262*

Dan G. Bachor

John O. Anderson

John Walsh

and

Walter Muir
University of Victoria

Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency

The last decade has witnessed considerable change in the ways teachers assess students in the classroom. The once ubiquitous classroom test has been augmented with a variety of emerging assessment techniques ranging from anecdotal observation to portfolios. Teachers' knowledge of how to evaluate the use of these new classroom-based assessment tools has not always paralleled their introduction. In this article we offer three criteria as guidelines for evaluating these classroom-based assessment techniques, which concern the representativeness, accuracy, and consistency of the information that is collected, interpreted, and reported.

La dernière décennie apporta des changements considérables dans la façon que les enseignant(e)s évaluent leurs élèves en salle de classe. On a ajouté aux tests administrés partout dans toutes les salles de classe, une variété de techniques d'évaluation à partir de l'observation anecdotale jusqu'à l'utilisation de "portfolios" descriptifs. Les connaissances qu'ont les enseignant(e)s à savoir évaluer l'utilité de ces nouveaux outils d'évaluation ne sont pas toujours allées de paire avec l'usage de ces nouveaux outils. Cet article offre trois critères comme guides d'évaluation pour ces techniques d'évaluation en salle de classe et démontre comment l'information collectionnée, interprétée et rapportée est représentative, juste, et constante.

Introduction

In the context of classroom assessment practices, there is an increasing use of a wide array of information collection procedures such as observation and portfolios (Bachor & Anderson, 1993). Although practices vary across grade

Dan Bachor is a professor in the Department of Psychological Foundations. His research interests include classroom assessment procedures as well as assessment and instructional strategies for special education.
John Anderson is an associate professor in the Department of Psychological Foundations. His research interests include classroom assessment practices and large-scale assessment.
Walter Muir is a professor in the Department of Psychological Foundations. His research interests, in addition to educational measurement and evaluation, include computer applications to education.
John Walsh is an assistant professor in the Department of Psychological Foundations. His research interests concern self-regulated learning in learning disabled students, cognitive strategy instruction, and assessing student learning.

levels and teachers, for the most part classroom assessment is not a discrete, single-instance group event such as the test. Classroom tests, however, do remain a component of teachers' assessment arsenal, the most common examples being weekly spelling tests for younger children and unit tests for older children. Other more formal tests are rare events and are removed for the most part from classroom activities. When they are given, such measures usually are restricted to a specialized purpose. Examples include administering individualized measures to make placement decisions for students with special needs, or requiring that all grade 12 students take graduation examinations.

Emergent forms of student assessment such as observation and portfolios are considered by teachers to be authentic evaluations of learning outcomes in that they are closely representative of the kinds of ongoing instructional activities in classroom learning (Bachor & Anderson, 1993; Nicholson & Anderson, 1993). Tests, on the other hand, are falling out of favor with teachers because they are seen to be too far removed or decontextualized from ongoing instructional activities (Wilson, 1990). This viewpoint is held by almost all teachers in kindergarten to grade 3 and by a growing number of teachers in grades 4 to 7 (Bachor & Anderson, 1993).

Reporting assessment information also is changing across grade levels. There is a shift from reporting highly condensed symbolic evaluation information in the form of letter grades to more descriptive, textual, or oral reports. Even where letter grades are still reported, they have been supplemented with comments that have tended to become more extensive.

Further, it is apparent that evaluating student achievement remains a common, important, and highly visible component of schooling. One aspect of classroom assessment practice that has received less attention than the particular format of data collection or reporting is that the practices of classroom assessment should themselves be consistently and continuously monitored and evaluated. The purpose of this article is to describe three criteria that can serve to provide a functional and feasible framework for evaluating many of the emergent classroom assessment practices. In meeting this purpose, we do not describe the full range of assessment procedures employed in the classroom. Instead, we restrict our examination to discussing two information collection techniques commonly utilized by teachers of primary (grades 1-3) and intermediate (grades 4-7) children: observation and portfolios (Bachor & Anderson, 1993).

Evaluating Classroom Assessment Practices

Traditionally, the evaluation of student assessment practices has centered on evaluating test quality in terms of validity and reliability. The focus of attention was on one test at a time and on each test administration. In other words, each instance of assessment activity was evaluated. This was generally done in conjunction with formal and often large-scale test administrations. Recently, validity in relation to the evaluation of student achievement has received considerable attention in the educational measurement literature (Messick, 1989; Moss, 1992; Shepard, 1993). In Messick's seminal chapter on validity (1989) and the more accessible chapter of Shepard (1993) the conceptualization of validity has been broadened to include not only the relationship of test items to the underlying traits being tapped, but also to consider the use of the

information generated by the test. Reliability generally has been equated with homogeneity of item performance within the item set composing the test. In addition, reliability is related to consistency of test results given variations in markers, the context of test administration, and in test questions (format and/or topics). The assessment of tests on the basis of validity and reliability would generally be done by measurement specialists associated with the administration and scoring of the test(s) in question.

In contrast, as noted above, classroom assessment practices consist of the melding of different information collection procedures developed and used by teachers that result in a series of data clusters for each student. Any guidelines that might be offered to aid teachers in evaluating their assessment practices must attend to the growing complexity of contemporary classroom assessment and be sufficiently straightforward to be worthwhile to classroom teachers.

Evaluation assessment practices can take the form of an appraisal of the discrepancy between intention and practice: what was intended to be assessed versus what was actually assessed. In the classroom context at least, there also must be a consideration of the quality and extent to which intention has been determined and articulated. That is, the intention must be clearly stated in terms of the established goals of instruction as translated into classroom practice. At least three criteria could be used to guide this evaluation of assessment practice. First, the extent to which assessment practices are based on tasks and activities that are representative of instructional intentions must be determined and then translated into specific information collection, interpretation, evaluation, and reporting procedures. Second, the accuracy of the assessment practices must be addressed. Third, the issue of consistency of these practices over time and across students warrants attention. A basic assumption underlying this discussion of the evaluation of assessment practice is that the classroom instruction is goal-directed and that these goals can be clearly known and described. Further, students will be assessed on the basis of the completion of some task such as writing an essay, responding to a multiple-choice test item, conducting a laboratory exercise, or demonstrating a physical skill such as dribbling a basketball. Finally, it must be noted that any change in the purpose of assessment should result in a revisitation of the specific application of the criteria of representativeness, accuracy, and consistency to any assessment task.

Representativeness

Assessment practice is representative to the extent that the task the student is to complete is clearly derived in a meaningful sense from the intended learning (the instructional) goals. For example, consider the goal of assessing problem solving. If the task may be successfully completed through the recall of information (such as remembering 5+6=11) as opposed to thinking critically about it, then the task would not be representative. A task that caused students to engage intentionally in critical thinking would be representative: 5+6 would be representative of problem solving if students did not know the answer. Instead, they reached the answer by problem solving by noting that 5+5=10 and that 6 is one more than 5, therefore 15+6=11.

Representativeness, then, refers to the extent to which the assessment information characterizes the learning agenda under consideration, where the

learning agenda includes both the broadly stated goals and expectations for students at a particular age-grade level and the more specific instructional targets set in any classroom. Wiggins (1989) may have had representation in mind in describing the concept of authentic assessment as an attempt to shift the focus of assessment practice toward collecting information on student performance that is more directly representative of the kinds of achievement that are valued.

In establishing a representative task in an assessment sense the learning goal has to be identified and described. The description should include as components both the process and product aspects of student achievement of the learning goal. In regard to process, attention should be given to the cognitive processes students use in the assessment task or activity. To illustrate, if multiple-choice questions were selected as an assessment device, the primary focus probably would be on the obtained answer, the product. An assessor might be able to infer how an answer was reached if there were a sufficient number of examples of any one type of answer. When observation is selected as an assessment tool, the tendency is to shift more to noting process.

Finally, when think-aloud protocols are selected, the goal is to unfold how someone is thinking. Thus with a shift in focus the emphasis in assessment can change from either product to process or from process to product.

The first step in selecting a representative task is the identification of valued learning outcomes. These are described in a manner that relates to observable activities or products and that captures the intent regardless of how complex or latent the outcome or goal. In the British Columbia context a major source of general descriptions of learning goals are Ministry program or curriculum documents (BC Ministry of Education, 1990a; 1990b; 1992). Another step in constructing or describing a representative set of activities is related more directly to assessment practices, that is, the development of evaluation procedures that capture both the process and the outcome.

Two examples are illustrative of these steps. In the first the initial task would be identifying and describing a learning outcome such as *communication of direct experience* in a manner that captures the intent and conveys the goal in an instructionally meaningful manner. This description then would be representative and would lead to identification of relevant indicators of achievement. The next step could involve asking students to compose an account of an experience they have had. This probably would be a more representative means of assessing writing ability than using a more indirect measurement format, such as giving a series of multiple-choice questions to check any student's understanding of the requirements of composition or determining if the child knows how to construct a sentence. As a second example, establishing an observation schedule that would result in samples of student behavior taken over time and across types of learning outcome is more representative of the full range of each student's repertoire of behaviors than completing a behavior checklist that calls for a composite picture based on recollections of past school-related activities.

Another aspect of developing representative assessment tasks concerns the extent to which such tasks reveal the cognitive activities of students. Assessment should provide opportunities for students to leave evidence of the

strategies and thoughts that they use of produce assessment products or other achievement-related activities. Just as the products and activities must be reflective of the content objectives of the curriculum, so too must assessment represent the cognitive strategy objectives that form the broad, but important goals of schooling. It is also important to recognize that these cognitive strategies will be utilized in situations that call for problem solving in an affective-laden situation (e.g., how do I avoid fighting when I am angry?) or when psychomotor skills are needed (e.g., how can I use imagery to relax before playing an important football game?) as well as in more obvious cognitive activities such as learning mathematics.

Developing assessment techniques that validly represent student cognition in addition to the products of such cognition is most difficult in the applied setting of the classroom (Marx, Winne, & Walsh, 1985; Royer, Cisero, & Carlo, 1993; Weinstein & Meyer, 1991). Certainly the precision of the cognitive psychology lab cannot be expected. However, there is emerging evidence that students can be trained to produce veridical information about cognitive strategies or plans that they use to accomplish classroom products.

A good example of this form of assessment is found in Swing and Peterson's (1988) study of mathematics problem solving. In this investigation students were trained to solve problems using adjunct questions to guide their problem solving tactics. Swing and Peterson assessed students' strategic use of these prompts by asking them to draw lines linking adjunct questions to appropriate parts of the problem. Of particular importance for the purposes of our discussion is that the students in this example left evidence of their strategic thinking as they solved problems. In this way, Swing and Peterson assessed both the products of student learning, that is, the completed problem solution, and the strategic thinking that produced the products, that is, students' use of adjunct questions.

The assessment of student cognition is not limited to the cognitive strategies of plans that students employ when responding to classroom work. Students can demonstrate their declarative knowledge in ways that reveal both the extent and organization of their knowledge. For example, students can group cards with concept words in order of close association. Clusters of closely associated concepts can then be arranged in superordinate and subordinate relationships. In this way students may reveal some of the changes in the structure of their knowledge as they become more expert in a particular knowledge domain (see Royer et al., 1993, for a more extensive discussion of the assessment of knowledge representation).

Accuracy

Accuracy refers to the level of error or distortion in the information generated in assessment practice. Accuracy concerns include the extent to which the intended data collection procedures were actually implemented. For example, say students were being assessed in terms of critical thinking by being asked to develop solutions for ambiguously described problems. The idea is to have students demonstrate cognitive flexibility in defining the problem and developing a number of plausible solutions (assume a representative task for critical thinking for this example). Yet, if in administering the assessment task either in the wording of the assignment or in teacher comments, students are

led to believe they are to derive single, best solutions for a common problem, then accuracy problems arise. The information generated by this assessment practice does not relate to cognitive flexibility per se but rather to identifying a standard problem and deriving a single best solution, and so is inaccurate to a degree. A more accurate assessment practice would result in students developing problem definitions and a range of solutions.

Another accuracy concern is the degree to which assessment practices in the classroom sample the characteristic performance of a given student. As an example of a situation with accuracy concerns, consider the classic test anxiety case. If a student is being assessed in the area of, say, critical thinking and perceives the situation to be one of high anxiety, then not only is thinking critically entering into the assessment information, but also the student's ability to cope with the anxiety. The information then is distorted; it is not accurately reflective of critical thinking.

Accuracy, then, refers to the veracity of the information collected and precision of any subsequent manipulations. It is important to note that accuracy is subsumed under representativeness; that is, each piece (or set) of collected information should be an errorless estimate of the larger assessment target (e.g., can the student add any combination of single-digit numbers?). There will be variations in the information collected to ensure it is representative; thus any variation in student performance or product should be directly related to underlying achievement for the two senses of learning what and for learning how. Student responses, activities, or products should be genuine expressions of achievement of learning goals of interest. For example, in responding to a multiple-choice item on a test of problem solving skills, a student should solve a problem in order to respond rather than simply recall a correct answer. In other words, the student is doing the activities considered to be essential indicators of the achievement of interest. In observing student communication, for example, an accuracy consideration would be seeing two students in oral communication as cooperative dialogue if that is what is going on, or as a threat to obtain free tickets to the new Sylvester Stallone movie if that is what is going on. Thus accuracy includes a determination that there is a match between estimate and target. In addition, accuracy is important in a second sense. Whenever information is collected, it should be recorded exactly, precisely, and unambiguously.

Accuracy considerations, then, follow from establishing a representative mapping of achievement related to tasks and activities. Each component of classroom assessment should be based on a solid information base that is as free from errors as possible in information collection, recording, interpretation, and reporting. For example, if the assessment question is centered on *What is typical student performance?* inaccurate information can result when only best products are selected, as they provide an incomplete and therefore inaccurate representation of the range of work typically completed. Products that might be more typical or that suggest the need for improvement will be missed if only best effort is used when collecting information (Bachor, 1993).

Consistency

Anderson and Bachor (1992) proposed that assessment procedures should be used systematically and consistently so that each student is evaluated for a

given learning outcome on the same basis as any other student. In other words, given the same underlying level(s) of goal attainment, the same kinds of evaluation will be made. In working toward this end, there should be consistency in the quality of procedures for information collection, interpretation, and reporting from one educational area to another and from one student to another *when the goal of assessment is to compare individuals or groups of students against some established criteria.*

Further, each area of schooling should be evaluated with similar standards of information collection and interpretation. For example, the quality and consistency of information collected to evaluate reading attainment should be similar to the quality and consistency of the information collected to evaluate attainment in mathematics or the goal area of physical development.

Several teachers (Bachor & Anderson, 1993) have argued against the use of consistent procedures, claiming that using the same type of procedures across students can result in unfair practices. The rationale for this assertion is that not all students can display what they are learning in the same manner. For example, some students may communicate effectively in writing, whereas others may be more effective visually or orally. As another example, setting a short-answer and essay unit test is unfair if the classroom contains a child who is physically handicapped to the extent that he or she is unable to compose written answers or unable to write at the pace required when taking such a test. Thus any consideration of consistency should be tempered to ensure that any adopted information collection procedure is feasible for all students. However, variations in data collection format can significantly affect the nature of the information collected and the meaningfulness of the conclusions reached.

Generally, however, adopting consistent procedures is suitable, and maintaining consistency increases the probability that representative and accurate estimates are obtained. That is, the goal is to apply the same standards of accuracy and representativeness for each student. An example of a consistency consideration would be that similar observations are being made in each of the observation sessions for that individual student, so that the obtained information can be used to judge if students have met widely held expectations for learning outcomes. Variation in the observed performance is due to variation in the levels of student achievement, not to variation in the methods of data collection.

In order to consider the utility of the criteria of representativeness, accuracy, and consistency in the evaluation of classroom assessment practice, two currently popular forms of assessment practice, observation and portfolios, are described. The use of these practices is discussed in relation to the three criteria discussed above: representativeness, accuracy, and consistency.

Information Collection: Observation and Portfolios

Observation and portfolios are two means of collecting information about student achievement that, as noted in the introduction, are considered authentic forms of classroom assessment practice in that the information collected and the way it is collected are considered to be fully compatible with the kinds of activities students do in classroom learning. In this sense observation and

portfolios could be viewed as formats well adapted to enhance the representativeness of assessment practice.

Observation

Observation typically is used to collect information about the behavior of students (e.g., the extent to which they cooperate or keep the classroom rules) and to note their activities as they learn (e.g., the degree to which they exchange ideas or present a viewpoint). It appears that the use of observation as an assessment practice is increasing in the elementary classroom (Bachor & Anderson, 1990; Board of Education of the City of Etobicoke Writing Committee, 1987; British Columbia Ministry of Education, 1989, 1990a, 1990b; Saskatchewan Education, 1991) and its use is relatively unstructured (Nicholson & Anderson, 1993). For example, the BC Ministry of Education (1989) introduced the use of observation as an assessment procedure in its draft version of the *Primary Program*:

> Observation is an important method of evaluating a variety of learning that cannot be assessed in other ways. For example, interpersonal behaviour, feelings, performance, thinking processes while solving problems, curiosity, and creativity are best noted through observation. Observation can help a teacher determine a pupil's knowledge of content, degree of skill development, and ability to apply skills. By observing pupils systematically in the natural setting of the classroom, the teacher can identify each child's unique interests, personality, learning style, strengths, differences, and difficulties. The teacher can use this information to plan programs that best meet the needs of individual pupils in the class. (p. 11.1)

We argue that using observation in the manner recommended above should be accompanied by close attention to the three criteria of representativeness, accuracy, and consistency.

The use of observation, as with every classroom assessment practice, should require that the teacher determine and articulate the goals of learning and the kinds of student performance or behavior associated with the achievement of these goals. In addition, observation requires that the teacher determine the situation(s) in which this performance is likely to occur.

As noted, a main rationale for the use of observation is that it is an authentic form of assessment because the kinds of things students are doing are directly related to learning activities, and hence the observational techniques are not decontextualized. The practice can be viewed as being representative in that the students' performances are closely related to learning outcomes.

Another representativeness consideration is the sampling of student activities. Sufficient numbers of observations each of adequate duration should be made in order to ensure that the activities observed were a representative sampling of the student's performance. What is sufficient would depend on the nature of the activities under consideration in that simple and relatively stable activities such as any discrete achievements in building vocabulary or increasing computational skills would require fewer observations than more complex activities such as critical thinking or more latent characteristics such as self-esteem or creativity. In terms of duration, an observation period would have to be long enough to allow for the targeted performance to occur, and for

the performance to be completed or at least enough seen that the gist of the performance is captured in the observation.

Accuracy in observation-based assessment concerns the veracity of the observations themselves and the procedures used to record them. Veracity of observations would be difficult to check in a classroom situation. In a more controlled situation multiple observers and perhaps videorecording events could be used to investigate this aspect of accuracy. Nevertheless, in classroom situations the veracity of observations can be enhanced by describing the foci of observations in low inference terms. This minimizes the extent to which observations are evaluated as they are being made. When possible and feasible, it is also helpful to develop and use a structured observation guide.

The consistency of observation procedures is also important. Observational procedures should be applied unvaryingly across students and for the same student across time. The first aspect of observational consistency is directed at ensuring students are observed in a similar manner and that approximately the same number of behavior samples are gathered for each student. In the second, consistency refers to collecting a similar number of samples across activities around the same period of time so that, as much as possible, each activity is portrayed in a parallel manner.

Portfolios

Portfolios are collections of samples of student work that are intended to represent the achievement of particular learning goals. Portfolios are described by the BC Ministry of Education (1992) as being collections

> of student work, gathered by the teacher or student over time, provide a rich array of information and evidence of learning ... When students and teachers develop portfolios that consist of judicious sampling of work done, however, collections become an integral part of learning. It is the substance of the portfolio that is important, not the form ... A portfolio must contain evidence of reflection and self-evaluation as this represents the student's efforts to learn about learning. Reflection and self-evaluation offer a concrete way for students to learn to value their own work, and by extension, to value themselves as learners. (pp. 106-107)

The portfolio generally includes work completed over a period of time and may include repeated drafts of the same piece of work in order to demonstrate the development (improvement) the student has achieved over time. The selection of materials to include in the portfolio generally is done by the student, either independently or with the teacher's assistance in judging whether the selection should be included in the portfolio. As a result of the student's active involvement in building a portfolio, and in contrast to observation and many other forms of assessment practice, in order to make meaningful selection and later evaluative-review decisions, the student should be aware of the learning goals germane to the relevant programs of study. Further, the student must comprehend the purpose of the portfolio. In other words, the teacher is not alone in the control of the assessment practice; the student is not simply responding to discrete items or test questions that have been developed or acquired by the teacher; rather, the student is responsible for assembling a collection of work completed during instruction and learning

that he or she considers to be representative of achievement of the goals of instruction.

The student is also responsible for the development or structuring of the assessment. This development would also involve collaboration between teacher and students on the goals of learning, the criteria for selection of work samples, presentation criteria and formats, and methods and criteria for evaluating the material. Each of the steps of collaboration must be considered in terms of representativeness, accuracy, and consistency.

Representative considerations for portfolios are similar to those for observation: work samples collected should be representative of the kind of performance associated with the achievement of learning goals; and there should be a sufficient number of these work samples taken over the time period under consideration. This requirement should be met by both the student and the teacher. This may require significant collaboration of the teacher and the student to establish goals and relevant (to the individual student and to the class in general) indicators of achievement. The purpose of compiling any portfolio also must be clearly understood by students and the teacher because this purpose will affect the number of samples included and the time period sampled. In addition, the range of material included will need to be considered. If the purpose is to present the full array of activities the student engages in, then work in progress, typical efforts, and best work should be sought. Conversely, if the intent is to illustrate the students' exemplary efforts to show to parents, then only the best material should be sampled. Further, the range of material needs to be considered from the perspective of the students' current capabilities so that written products, audiotapes, and videotapes of activities could be included to ensure that the portfolio includes options to allow the illustration of kind of performance matched with the achievement of learning goals.

Accuracy would focus on whether the targeted data (work samples) are in fact collected. The work samples would have to be the genuine article in the sense the student did indeed produce the work in an instructionally relevant context. For example, issues such as plagiarism and external assistance may come into play, but this would probably not be a major source of problems with the portfolio in the context of classroom assessment. Such issues could be more important where computers are being used, especially when databases are available where material can be readily exchanged. In this case, establishing who wrote what will become difficult. The net result will be that to determine accuracy, that is, a genuine article, could be extremely difficult if not impossible.

Perhaps a greater threat to the accuracy of work samples is found when classroom tasks are completed in groups. Several researchers (Corno & Mandinach, 1983; Marx & Walsh, 1988; Rohrkemper & Corno, 1988) have found that group work provides opportunities for some students to disengage from learning. For example, students may adopt purely management roles in group work. These students, called resource managers by Corno and her colleagues, are adept at gathering and presenting information but leave much of the cognitively engaging work of integrating information to others. In a similar vein, other students may rely almost exclusively on the efforts of others. These

recipient learners (or perhaps nonlearners) acquire and transform information only to glean the gist of what was accomplished in the group. In still other cases, disengagement from participating in the group activity can mask lack of confidence or competence that may be missed by the teacher when group work is prevalent.

Whether students adopt a task management, recipient learner, or other approach to group work is perhaps of lesser importance than is the way that the accuracy of work samples may suffer as a result of different styles of task engagement. Students can produce group work samples that may give the appearance of accurately reflecting learning, but in fact reflect little genuine cognitive effort. Without taking more direct data on the thinking of students as they complete group tasks, the accuracy of the meaning of the work samples may well be reduced.

Several techniques may be used to assess whether work samples completed by students accurately reflect their thinking. Paralleling closely the stimulated recall procedures used in the psychology laboratory, teachers can present work samples to students and ask them to recall the procedures, strategies, and general thinking that created the product. Consider, for example an assessment interview that might be conducted following instruction on poetic techniques. During the interview, a student might be asked to recite a poem she or he has just written. At various points during the recitation, the teacher might interrupt asking why a particular metaphor was used, why onomatopoeia was selected, or why iambic pentameter verse was chosen. By being questioned while the poetry is recited, students' recollection of their thoughts during the initial writing of the poetry are enhanced (Marx et al., 1985). In this way students' reports of their thinking can be stimulated by presenting their academic work during assessment interviews.

Alternatively, students may be instructed to leave evidence as to the way a task was completed during their solution of the problem. Annotated outlines of essays, rough notes of group presentation, and detailed calculations in algebraic problem solving are a few examples of myriad ways students can leave traces of their thinking (Marx et al., 1985). These intermediate academic products can, of course, also be used as the foci of assessment interviews as discussed above.

Consistency considerations can be problematic in the use of portfolios. First, if students are involved in the selection of work samples representative of given learning goals, there must be a common and consistent understanding of what these goals are in a substantive manner between teacher and student and across all participating students. Second, there has to be a common and consistent understanding of the purpose of the portfolio itself, particularly between teacher and student.

Evaluating Evidence and Reporting

The next logical steps in assessing student achievement are to interpret the evidence collected and to report these interpretations. The teacher must make sense of the observations or the portfolio if such information is to inform instruction and aid in making judgments about the achievement of students. In regard to these interpretations, evaluations, and reports, the evidence interpreted should be representative of that collected, and in turn should be repre-

sentative of student performance closely related to learning outcomes. The evaluative judgements should be accurate and consistent. Reporting the evidence and the evaluations should be representative of the student's achievements and the field of learning outcomes, accurate descriptions of evaluations and supporting evidence, and consistent across students in terms of coverage and quality.

Traditionally, teachers have tended to collect information most intensively around reporting periods rather than systematically across the course of the school year. Their typical practice has been to compile any current information and/or to invoke their memories about a student's performance. Following this preparation period, report cards were written and sent home. Parent-teacher interviews followed if they were deemed necessary.

As teachers change evaluative practices, at least two components of the assessment process require clarification. First, the common interpretive practice appears to be to select and report rather than to evaluate the gathered evidence; that is, judgments made by teachers are often quite subjective. Thus we argue that some guidelines are necessary to ensure that the evidence presented is representative (meaningful), accurate, and consistent (fair). Similarly, the rationale and formats of reporting need to be examined to ensure that reporting practices meet the same criteria.

Interpretation

Interpretation of the data collected by assessment practices, such as through observation or portfolio use, should involve some checks of data quality and will general involve substantial reduction of collected information into more manageable aggregates of evidence. To check the quality of the collected data the teacher should evaluate the representativeness, accuracy, and consistency of the gathered information before making any judgments about student achievement. We argue that most of these checks should be done before and during collection of the information. Before examining the resultant data, however, these considerations should be revisited. Further, steps should be taken to ensure that the overall purpose of conducting assessment is maintained.

The reduction of assessment data should be conducted in a way that maintains close linkage to the purpose of the assessment and to the learning outcomes under consideration. To do this will require planning on the part of the teacher. Bachor (1993) recommended that the teacher should begin to interpret information by asking about the extent of evidence required to make judgments about student achievement of a particular learning outcome. This would be directly related to the guidelines we recommend for information collection discussed above. This is recommended in the context of interpreting information as a cautionary step prior to making any substantive decisions about students.

The kind of interpretation teachers make will be one of location: given this information about student performance or product in relation to learning goals, where is the student located in terms of achievement? (Bachor & Anderson, 1993). The issue here is to establish where the students are in relation to established learning goals. To do this student performance is compared with expectations. The expectations that are the basis of this comparison are the

performances expected of achieving and nonachieving students. To enhance the consistency of interpretation, these expectations should be explicitly considered and described by the teacher. As an example, if the learning goal for a class is to *effectively analyze news reports in terms of fact or opinion*, the kinds and qualities of products (in this example written reports) that would indicate achievement or nonachievement should be fully described by the teacher and known by the students for pedagogical reasons. This would facilitate consistent teacher decisions about the achievement status of the students generating the assessment date. It should be noted that the development of a full range of expectation descriptions would be ideal in the maintenance of consistency.

To aid in locating students' accomplishments in an achievement space or on a continuum, numerous scales and frameworks might be used. The use of a well-developed scale or framework could also assist in maintaining consistent interpretation. If teachers have a consistent structure on which to place or locate various levels of students' products or performances, teachers' judgments are likely to be more consistent than if the scale or framework is constantly being created anew for each student with changing meaning associated with similar achievement labels. The most common scale is the letter grade scale. Another, recommended by Bachor (1993), is a framework based on four clusters of performance, where any information relevant to the purpose of classroom assessment could be described under the following labels: capable (can do above level of expectation, competent), characteristic (typical of what the student can do), acquiring (what the student is learning), and unknown (at this time the student is unable to do X).

Regardless of the scale used in reporting, the three recommended criteria must be addressed to determine if students have been located accurately and consistently across the established goals. Possible interpretive errors in making this decision are addressed below.

Possible interpretive errors. Nuttall and Goldstein (1984) list four errors in judgment a teacher can make when reading assessment information: (a) halo effects; (b) leniency or severity effects; (c) central tendency errors, which are essentially errors associated with our criterion of accuracy; and (d) sampling information from a restricted range, which is an error associated with representativeness.

Halo effects occur when teachers' judgments are based on belief rather than performance; that is, teachers judge what they expect overall performance to be rather than examining the activities actually completed. In making errors of leniency or severity, the conclusions reached are not warranted by the collected information. The judgment made is that the performance is poorer or better than is actually the case. Leniency effects can occur when a teacher wants a particular student to do well, accepting an assignment or other work as being "good enough" rather than meeting the desired standard. Severity effects are found when too harsh a judgment is made. The third class of error suggested by Nuttall and Goldstein (1984) is the central tendency error. In this case only typical information is collected while extreme performance is ignored. As a consequence, the full range of potential performance is unknown to the teacher, which would be an error if the purpose of assessment were to evaluate the full range of student performance. The last type of error is in

sampling information from a restricted range of possible performances. The restriction is not in terms of the range of performance of a given task by students as in central tendency error, but rather in terms of the kinds of performances considered. The performances observed would reflect a restricted range of learning and teaching conditions. For example, a student who is being judged on mathematics achievement may be classified incorrectly if the work being considered represents performance relying primarily on the student's reading ability (Bachor, 1993).

Reporting

One purpose of classroom assessment (and a legal requirement) is to report to parents. To some degree teachers have little choice as to the number, format, and timing of the issued report, as such matters typically are prescribed. However, teachers still have to address a number of practical questions when they report: How complete is the presented report to be (the extent of coverage of each goal for any student and to a lesser degree the number of goals to be addressed)? Should the report be supplemented by a conference (assuming one is not required), or if only an oral report is needed should it be written instead? If a conference is to be held with parents, who should participate? To some degree teachers in any school also are varying timing of reporting delivery (Bachor & Anderson, 1993). In these cases teachers are asking about the timing of the delivery of any report; should it be before, after, or during a conference with parents?

Reports on student achievement should be representative in terms of representing the range of goals reported and the range of student achievements in the goal areas. In terms of goal coverage, the report should not be confined to a restricted range of goal areas: all relevant and important goal areas should be targeted. In terms of achievement coverage, the full spectrum of student achievements should be reported. As an example of a restricted range of performance being reported, in the early grades in British Columbia (up to about grade 4) there was an emphasis on reporting only those student achievements that could be perceived as a positive reflection of what the student had accomplished. These were the so-called can-do statements of student achievement. In effect, a restricted range of goal areas and student achievements was reported for many students (Bachor & Anderson, 1993). In some cases, can-do reports were summaries of the student's learning history, rather than a reflection of what was being learned in the current year. In neither case would the resultant report be representative of the established classroom goals.

Accuracy would essentially focus on the truthfulness of the reports. The accuracy would be determined by the preceding steps of information collection, interpretation, and evaluation. Given that these preceding steps were representative, accurate, and consistent, the chance of the report distorting the student's current achievement should be minimal.

Written reports vary from comments alone (termed an anecdotal report) to combinations of commentary and letter grades. An anecdotal report is a narrative on student achievement but may also include summary descriptions and evaluations of student abilities and characteristics. The assignment of evaluations and comments to students is based mainly on teachers' tacit understanding of students (Bachor & Anderson, 1993). Similarly, the basis of the

assignment and the form of letter grades varies greatly, making interpretation of this kind of report more difficult for parents and students. So, certainly, the meaning of symbols, phrases, and other characterizations of student achievement has to be fully explained to parents and used consistently across students and across reporting periods to enhance the utility and meaningfulness of classroom assessment practice. An exception will occur when students with special needs and individualized programs are added to any class. In such cases, there might be some degree of inconsistency in the reporting format and content.

Several new forms of reporting are emerging in popular practice in British Columbia schools. Among these, student self-evaluation and various forms of conferencing are prominent. These two reporting procedures may take a variety of forms; nevertheless, the criteria of representativeness, accuracy, and consistency are unvaryingly important across these forms. To discuss these criteria in relation to reporting is a topic for future work.

Summary

We argue that practitioners need to evaluate carefully classroom assessment and reporting practices. The quality of these practices is not well known, although the importance and presence of classroom assessment practice is considerable. In order to assist practitioners in their evaluation of their classroom assessment practice, we suggest the use of three criteria: representativeness, accuracy, and consistency.

References

Anderson, J., & Bachor, D. (1992). What should a classroom testing program look like? The functional factors of an assessment program in primary classrooms. In D.J. Bateson (Ed.), *Classroom testing in Canada* (pp. 59-68). Vancouver, BC: Center for Applied Studies in Evaluation, University of British Columbia.

Bachor, D. (1993). The changing face of assessment. In J. Andrews & J. Lupart (Eds.), *The inclusive classroom: Educating exceptional children* (pp. 293-329). Toronto, ON: Nelson Canada.

Bachor, D., & Anderson, J.O. (1990). *Assessment practices in the primary program: Description of observed practices, functional factors, and recommendations as to some general principles: Final report*. Victoria, BC: Ministry of Education.

Bachor, D.G., & Anderson, J.O. (1993). *Assessment practices in the elementary classroom: Perspectives of stakeholders*. Victoria, BC: Ministry of Education.

Board of Education of the City of Etobicoke Writing Committee. (1987). *Making the grade: Evaluating student progress*. Scarborough, ON: Prentice-Hall.

British Columbia Ministry of Education. (1989). *The primary program: Response draft*. Victoria, BC: Author.

British Columbia Ministry of Education. (1990a). *The primary program: Foundation document*. Victoria, BC: Author.

British Columbia Ministry of Education. (1990b). *The primary program: Resource document*. Victoria, BC: Author.

British Columbia Ministry of Education. (1992). *Intermediate program: Draft, March 1992*. Victoria, BC: Author.

Corno, L., & Mandinach, E.B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18*, 88-108.

Marx, R.W., & Walsh, J. (1988). Learning from academic tasks. *Elementary School Journal, 88*, 207-219.

Marx, R.W., Winne, P.H., & Walsh, J. (1985). Studying student cognition during classroom learning. In M. Pressley & C.J. Brainerd (Eds.), *Cognitive learning and memory in children: Progress in cognitive development research* (pp. 181-203). New York: Springer Verlag.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*, 229-258.

Nicholson, D.J., & Anderson, J.O. (1993). A time and place for observations: Talking with primary teachers about classroom assessment. *Alberta Journal of Educational Research, 39*, 363-374.

Nuttall, D., & Goldstein, H. (1984). Profiles and graded tests: The technical issues. In J. Mortinare (Ed.), *Profiles in action* (pp. 5-150). London: Further Education Unit.

Rohrkemper, M., & Corno, L. (1988). Success and failure on classroom tasks: Adaptive learning and classroom teaching. *Elementary School Journal, 88*, 297-312.

Royer, J.M., Cisero, C.A., & Carlo, M.S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research, 63*, 201-243.

Saskatchewan Education. (1991). *Student evaluation: A teacher handbook*. Regina, SK: Author.

Shepard, L.A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Research in education: Vol. 19* (pp. 405-450). Washington, DC: American Educational Research Association.

Swing, S.R., & Peterson, P. (1988). Elaboration and integrative thought processes in mathematics learning. *Journal of Educational Psychology, 80*, 54-66.

Walsh, J. (1992). Promoting self-regulation in the learning disabled. *Exceptionality Education in Canada, 2*, 1-14.

Walsh, J. (1993). The promise and pitfalls of integrated strategy instruction. *Journal of Learning Disabilities, 26*, 438-442.

Weinstein, C.E., & Meyer, D.K. (1991). Implications of cognitive psychology for testing: Contributions from working in learning strategies. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition* (pp. 40-61). Englewood Cliffs, NJ: Prentice-Hall.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 703-713.

Wilson, R.J. (1990). Classroom processes in evaluating student achievement. *Alberta Journal of Educational Research, 36*, 4-17.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1*.
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
4-116 Education North
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1436



THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XL NUMBER 3 SEPTEMBER 1994

The Alberta Journal of Educational Research

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
University of Alberta

George Buck
University of Alberta

James Chapman
Massey University

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Linda LaRocque
University of Alberta

Stewart J.H. McCann
University College of Cape Breton

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Sharon Robertson
University of Calgary

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Graham Fishburne
Robert H. Short
Fern Snart

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Publication Services
4-116 Education Centre North
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0390

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XL, Number 3, September 1994

Editorial

Judy Cameron **263** Higher Education Under Attack in Alberta: Short-term Versus Long-term Consequences

Perspectives

Sara Stambaugh **265** Getting Educated

C. Del Litke **271** Implementing the Modified Four-day School Week

Articles

David Watt
Hetty Roessingh **283** ESL Dropout: The Myth of Educational Equity

Ardra Cole **297** Problems and Paradoxes in Beginning Teacher Support: Issues Concerning School Administrators

Paul T. Begley **319** Contemporary Images of School Leadership in Ontario, the Northwest Territories, and Western Australia

Genevieve M. Johnson **337** Undergraduate Student Attrition: A Comparison of Students Who Withdraw and Students Who Persist

B.E.J. Housego **355** How Prepared Were You To Teach? Beginning Teachers Assess Their Preparedness

John Connors
Robert Frender
Patricia Hughes-Fuller
Frank Symons **375** Book Reviews

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 263-264*

Editorial

Higher Education Under Attack in Alberta: Short-term Versus Long-term Consequences

We live in a time when many industrialized countries, including the United States and Japan, are concerned about the quality of education. Japan has a long history of focusing on and financing educational pursuits. In the United States, it is widely recognized that in order to be an international leader and to be competitive with other countries, the educational system must become a priority. In Alberta today, the view appears to be the opposite. In a time of financial difficulty, the government of Alberta led by Premier Ralph Klein has decided that in order to create a competitive business environment, education must be slashed. Although these cuts affect all areas of education, as a university professor my focus in this editorial is to describe the impact of such cuts to the university system.

Let's first consider the reasoning behind a policy that dictates a severe reduction in educational spending. The intention of the government is to make a lean, mean, and competitive economic environment where taxes are low and the province attracts business ventures and outside capital. Indeed, in the short run, by cutting education (and other public services) the government does save money. There is, however, a major hitch. In terms of higher education, class sizes increase, research declines, many of the best professors leave, new scholars are difficult to attract, and the overall morale of those who stay is low.

In accord with the government's goals to cut funds, there is a push to emphasize the training of job-related skills. This emphasis has, in my view, serious negative consequences. During a time of limited resources and staff, we in the university system are being asked to spend more time teaching the practical implications of our subject matter. Traditionally, and for good reasons, a university education has focused on analytical thinking, critical evaluation, writing skills, research competence, and so on. Many of these skills will no longer be developed in students who receive their degrees in Alberta. With three or more courses a week to teach and class sizes that often exceed 50 (more likely 100), it becomes impossible to foster these types of skills. Furthermore, professors who want to keep up with their field and do research will no longer have the time to do so.

Interestingly, these profound alterations are often touted as positive changes that will propel our higher education system into the next century. The implication is that these are forward thinking, positive new directions for higher education. The twist is, however, that in the call for change, the university goes from a dynamic system (a system that produces inquiry and research) to a static one (one that operates only on existing knowledge and resources). Rather than generate new knowledge, professors will parrot existing facts and they will do this only when those facts have an immediate practical payoff.

A good university education is one that prepares students for the future rather than the present. Students may read philosophy, wonder about superconductivity, analyze sentence structures, evaluate cultural practices, research 18th-century English authors, or investigate the navigation of homing pigeons. Such graduates may take a few months to fit into an existing business; some may not generate anything practical, and others may never increase capital in the province. A few, however, will make excellent business partners who have the skills to change when the business climate changes. Some graduates may do something akin to discovering insulin; many others will contribute to a province that is concerned with more than the immediate economy. In fact, it is these students who will (if they exist) take Alberta into the next century.

In the long run, by the time it becomes evident that a decimated university system has serious negative implications for a robust economy, enormous cost will be involved in recovering what has been lost. It is not simply a matter of changing policy and adding money in order to regain an excellent university. Professors will have gone, the university will have a poor reputation, students will have been poorly taught, and centers of excellence will no longer exist. In the short run expenditures will be reduced, but at an immense future cost to all citizens of the province.

Judy Cameron

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 265-269*

Perspectives

Getting Educated

Sara Stambaugh
University of Alberta

I decided to begin, like David Copperfield, with being born. Whether or not my birth gave me insight, I was born in a caul (again like David Copperfield) and delivered at midnight by, believe it or not, Dr. Foust. In spite of my obvious qualifications to be a seer, I won't present my visions of the future of higher education now that the president of the University of Alberta has put *Quality First* in a glossy publication that pays lip service to quality and uses double talk to undermine it. Instead, this brief chat will be a glimpse of what I used to think education was about in the days when professors of English and other educators still had clear ideals and standards. To explain some of the changes since my youth, I'll tell you about my own education and some of the people who helped me to get it.

To begin, I was a country kid born in the Depression, and both my parents were straight off the farm. That meant that my mother had to milk the cows and do the farm work before she went to school, as she fought to do through grade 11, when the local school stopped. My father was more typical and sat through, I think, five years in a one-room schoolhouse. Obviously, in those days education belonged to the rich or, perhaps, to city people who had access to a high school.

By the time we settled into southeastern Pennsylvania after the War and my turn came, high school was available, but college wasn't. As grade 12 had been for my mother, college (the American equivalent of university in Canada) was reserved for the rich, but I had my heart set on it.

I had two reasons, first, my terrible crush on Fred Myers, a beautiful and cultured young man in my class. I pined over him as my first love and ideal. Mooning over Fred didn't make life as a teenager any easier, but Fred gave me a standard and kept me from marrying Jakie Groff or some other country boy when I was 18. (Incidentally, years later I met Fred and his wife and children on a train to New York, and I thought to myself, "What fine taste you've always had, Sara.")

The second reason I wanted to go to college was my music teacher, Virginia Darnell. When I was in grade 7 the school music teacher, Mrs. Schwackhammer (we have wonderful names at home) thrust musical instruments at assorted students. She handed me the extra flute, taught me basic blowing, and an-

Sara Stambaugh is the author of two novels, *I Hear the Reaper's Song* (1984) and *The Sign of the Fox* (1991), and of a critical study, *The Witch and the Goddess in the Stories of Isak Dinesen* (1988). Since 1969 she has been a professor of English at the University of Alberta.

nounced that I needed a real teacher. My mother dug into the money she'd been scrimping to buy a new refrigerator, and I started to study with Virginia.

I think I can call Virginia Darnell a mid-century feminist, one of the sort Isak Dinesen refers to as disguising themselves as men, because the climate was so rugged for independent women in the '40s and '50s that they had to tough their way through a male world. Virginia wasn't an intellectual, but she'd toured North America playing bassoon with a women's woodwind ensemble.

When I met her, Virginia was sitting in on French classes at the high school across the road so that she could read French books and magazines, and she made regular trips to Philadelphia to visit friends in the Philadelphia Orchestra. Philadelphia was 60 miles away, a little over an hour by train, but for country people it was the end of the world. As an example, my best friend in high school lived outside Lititz, less than 10 miles from the county seat, but she'd only been to the city once.

Virginia passed me to a flute specialist when she'd done as much as she could to teach me, but she kept me under her wing. She introduced me to local chamber music groups, and she pulled strings so that I could spend my senior year at Lancaster Catholic High School and compete for college scholarships. Subsequently, she even arranged an audition for me with Marybelle Nissley, who offered me the chance to play third-chair flute in the Women's Air Force Band—just in case I wanted to follow in Virginia's footsteps.

I was at the Catholic high school only a year, allegedly to strengthen its music program, though clearly it needed more than I could give it. Still, my Catholic education was invaluable when I came to teach. Besides, during all the masses I had to sit through, the nuns let me read my Bible, and while everybody else was genuflecting, I was absorbing the Latin mass and reading the raciest parts of the Old Testament—both of which have stood me in good stead.

Thanks to Virginia, I won my scholarships and got to college, rather than to the Women's Air Force Band, the local state teachers' college, or a nearby hospital to learn nursing. Bright men could live at home and go to a local liberal arts college for men. I went to Philadelphia to Beaver College which, in spite of current connotations, I can't praise enough. It had a fine English department, and my second adopted mother, Margaret Green, was my freshman English teacher.

Margaret had the reputation of being the meanest, toughest teacher in the department. If she was, it was because she had high standards. She'd worked to get her education during the depression, and she had a profound respect for literature and for the language. PhDs were rare, and Margaret had an MA from Columbia, probably at least as sound a background as most current PhDs. Our chairman, Doris Fenton, did have a PhD, from the University of Pennsylvania. Word was that every year Bryn Mawr tried to hire her away from us.

Still, neither Doris Fenton nor Margaret Green published. Publishing then was for the rare insight or, especially, the comprehension evolved from a lifetime of study, and neither of my teachers would have been so presumptuous. Instead, both of them spent their summers studying and reading everything available about what they had to teach. Interestingly, I realize

now the context the two of them came out of. Both had degrees in classics, the *sanctum sanctorum* 19th-century men had previously reserved for themselves so that they could show their superiority. In other words, my teachers at Beaver were the real avant garde of the women's movement, never strident, always womanly, and invariably professing the highest of moral and intellectual standards.

Neither was a stimulating teacher as measured by the student evaluations now in favor. Margaret terrified her students, and Doris had a high, up-and-down voice just like Eleanor Roosevelt's. I remember sitting through some of her classes with my fingers carefully propping my eyelids open. But I wouldn't have passed my PhD examinations if Dr. Fenton hadn't meticulously presented Milton's classical sources (not in the most stimulating lectures) and, especially, explained Shakespeare's use of tumbling verse in the early comedies—pretty good, I'd say, for undergraduate courses. Both Doris and Margaret were genuine scholars who cared more about passing on knowledge than about entertaining their students. Their central concern was that the students in their classes develop a clear understanding of the works and of their intellectual context, a focus I see emphasized less and less in current English programs.

Thanks to them and my other teachers, by the time I graduated from Beaver I'd gained an appetite for studying more, and once again Virginia Darnell came to my rescue. She proclaimed (Virginia always proclaimed) that to avoid being parochial, I had to go to a different region and to a place with a major symphony orchestra, a measure, she said, of the cultural life. Besides, Virginia came from Minnesota and had a friend on the admission board at the university.

I can't speak for the quality of the department now, but when I was at the University of Minnesota the staff included Huntington Brown in Old English, Jacob Levenson in American literature, Leonard Unger in 17th century, G. Robert Stange in Victorian, Allen Tate in 20th century, and Samuel Holt Monk in the 18th. All my teachers were addressed as *Mr.*, partly because Allen Tate didn't have a PhD, partly to distinguish themselves from professors of education, and partly because they were trying so hard to imitate Harvard and Yale. There was one woman on staff.

Here it's worth mentioning the basic difference between British and American graduate studies (at least as they were in the '60s). Under the British system the emphasis was generally on a central figure or subject, with the assumption that the student would subsequently expand and figure out the rest. The American system, however, was the opposite. Students had to learn a fair amount about everything and only subsequently home in on a subject of concentration. As a result, graduate students at Minnesota took as many courses as they could, because they were tested on them before they could move on, and when they went into the world they had to be ready to teach any course in any area.

That was the reason I took a year in 18th-century literature with Samuel Holt Monk. He didn't adopt me, and from the sex of the students he did adopt, I suspect that he was a totally innocent male chauvinist in an age when southern gentlemen (and others) didn't know better. It doesn't matter, because Mr. Monk showed a generation of young women as well as men what it meant

to be a gentleman and a scholar. Students who went to Minnesota to study Shakespeare took Mr. Monk's course and, instantly infected by his love of his subject, shifted to the 18th century.

Students, of course, are nosey about teachers they like. Rumor had it that during World War II Mr. Monk had personally been responsible for saving an unnamed European cathedral city from allied bombing. Word also was that he'd stood helplessly on the shore and watched his wife and children drown in a motorboat accident. Whatever the truth of these stories, I can personally attest that Mr. Monk was a hypochondriac. Moreover, he took Samuel Johnson's injunctions against cant so much to heart that passing him in the hall and saying "How are you" invariably resulted in a list of symptoms and ailments 15 minutes long.

Nevertheless, Mr. Monk was a model to his students of what a professor should be. He began his lectures by writing on the blackboard all the current books and articles he'd drawn from or that were pertinent to the subject. However, like my other favorite teachers, he'd have failed current teaching evaluations. Mr. Monk stuttered. He got terrific stage fright every time he had to face a class. I well remember all of us leaning forward and trying as hard as we could to encourage him to get a word out, while he stood paralyzed and sputtering at the front of the room.

When he could speak, however, his love of his subject took over. And his love was contagious, especially because it was based on so profound a knowledge of the 18th century. Talking about Pope's Belinda, for example, Mr. Monk suddenly got shy about her toilet table, blushed (as a gentleman from Virginia should), and only hinted at the unmentionable ingredients used in 18th-century cosmetics. Another time, carried away by his subject, he inadvertently alluded to one of Alexander Pope's early sexual adventures. He stopped suddenly, stuttered to a halt—and gave his students even more reason to study the 18th century.

Sam Monk also had a meticulous eye and ear for the language. One of my friends who wrote his thesis with him told me that he once deliberately slipped in a participial phrase without a specific grammatical antecedent, as a test. Of course Mr. Monk caught it and lambasted Bill for sloppy writing.

The quality of Mr. Monk's own prose is clear from the section he contributed to *The Norton Anthology of English Literature*. Recently, rereading his introduction to the 18th century brought tears to my eyes, because it is so imbued with the spirit of the man, who, like my earlier teachers, taught me that the ideal of writing was to present a complex subject so lucidly that anyone could understand it. So far as I can tell, Mr. Monk's introduction to the 18th century is the only section not largely rewritten in Volume I of the latest edition of the *Norton*, I assume, because the current editor realized that Mr. Monk's prose and lucid explanations of, say, the workings of the heroic couplet couldn't be bettered.

Still, by the current standards used for university promotions, Mr. Monk would not do well. After all, he stuttered, he sometimes all but fell into paralysis in the middle of class, and, especially, he didn't publish enough. His reputation was built upon one brilliant book, *The Sublime*. Nevertheless, what

he wrote was important enough to shift perceptions about the 18th century, he was known and respected by leading experts across the English-speaking world, and he formed a generation of scholars.

I've told you who my models were and what they taught me, but in each case they taught one central lesson: that teachers teach ideals, which, at least in my own profession, seem nowadays to have dissipated into nebulous and contradictory theories about the function of the teacher, the purpose of language, and the very nature of English studies. Though I was born in a caul and should have second sight, I don't have a vision for the future. What I can see, however, isn't *Quality First* for our present and future students.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 271-281*

Perspectives

Implementing the Modified Four-day School Week

C. Del Litke
Lacombe Junior High School

On February 1, 1994 Lacombe Junior High School in Alberta, under the leadership of principal Don Pollock, implemented a pilot project of a modified four-day school week that extended from February 1994 to June 1994. Using existing county busing schedules and policy framework, the 500 students at Lacombe Junior High started their day five minutes earlier, ended their school day five minutes later, and had their noon hour shortened to 30 minutes. The outcome was a schedule that consisted of nine 40-minute classes as opposed to eight 40-minute classes. As a result, 81 classes were taught in a nine-day period as opposed to 80 classes previously taught in a 10-day period, which allowed both staff and students to take every second Friday off.

It had been five years since I left my master's program at the University of Alberta. I was ready to take on the world of educational administration, and I was armed with the research of my thesis on educational change. I would be a change agent; I had done my homework. I possessed the knowledge of extensive study of Fullan, Hord, Loucks, Hall, and others regarding change: "Change is a process, not an event." "Change is a personal decision" (Hord, 1987, pp. 93-94). Well, halfway through a public forum I had a revelation. Change may be a personal decision, but all these angry parents are an awfully tough *group*. The only personal part about this decision seems to be the attacks on our administrative team. I haven't slept well for weeks, my stomach is constantly in knots, and I jump three meters in the air whenever a phone rings. Hey, you writers, I think you forgot to warn me about a few things! I now realize why change happens so slowly in education. I didn't realize I had so many bosses.

The Initial Thought

It is appropriate to mention at this time that my principal is either extremely creative or nuts. When we get together and talk about the school, he is prone to come up with what he calls his "Scheme of the Week," which is often an idea far from the mainstream of educational thought and practice. Lately I had been lamenting about the difficulty of his schemes. I had been reading Senge's (1990) *The Fifth Discipline: The Art and Practice of the Learning Organization*, and I was struck by his concept of *systems thinking*. I pointed out that every time we had a good idea, we had to run over to the timetable and see if it would schedule or check with the busing schedule, or ... the problem as I saw it wasn't with the

Del Litke has 12 years of secondary school experience and has been Assistant Principal for five years. He completed his master's degree at the University of Alberta.

scheme; it was with the system, the timetable, the schedule. I didn't realize he was listening.

After a particularly contentious staff meeting, we again began to consider seriously the problem of noon supervision. It was apparent that a large majority (420+) of our student body of 500 was staying at lunch hour, causing enormous stress for supervisors, discipline problems for administration, and clean-up problems for our maintenance staff. Our school ran extremely well when the students were in classes; unfortunately, it was chaotic during noninstructional time. Again the system came into to question. How can we address our noon-hour problems? Our system locked us in with timetables, bus schedules, other schools using our facilities, and provincial requirements for instructional hours. Perhaps our answer was in examining the system itself. After all, extensive funding cuts by the Alberta government to education in the province of Alberta demanded that the education system be changed and that a basic education be defined. Preceding these cuts were a number of "Educational Round Tables" that were held throughout the province designed to solicit input from all educational stakeholders to determine which elements of education were basic and should continue to be funded and which elements of education were nonessential and should be earmarked for funding cuts. It was apparent in all three round tables that I attended that noon-hour activities and intramural programs were identified as frills, and that they should not be part of funding a basic education program when cuts were to come. Exploration of other ideas such as four-day weeks and year-round schooling were encouraged at the round tables to maximize facility use and minimize the costs for providing basic education. Hence when we began to examine the system, the seeds of the idea of the modified four-day week were sown.

After throwing the idea around, our administrative team agreed to give the four-day week concept some thought, which I thought meant the issue was dead. Often when we agree to give things thought we simply don't have time to reconsider the idea. However, then our principal came in with his latest Scheme of the Week: The Modified Four-day Workweek. During Christmas vacation we met to discuss the specifics of the proposal, and a meeting was scheduled with central office to discuss the proposal. Our outline was to receive central office permission, staff permission, then student and parent permission.

The original meeting with the superintendent and his deputy left us with a mixed reaction. We did not receive permission, nor were we rejected. Our principal was scheduled to make his pitch to the board of education at their next meeting. History dictated that this meant the project was finished. Our school board had not been in favor of any schedule changes in the past. It was also suggested at this meeting that if instructional time could be increased or prep time could be addressed, the proposal would be more saleable to the public. It was also suggested that we not go to the public with our proposal before the board meeting because it upset the decision makers to be the last people notified. This would prove to be a controversial decision later.

We decided to approach our teachers next with the idea because if they weren't in favor, there would be no use going to the board. Here is the time that

one of Fullan's principles on change reared its head: *characteristics at the district level* (Fullan, 1982, p. 56), specifically history of board-teacher relations. A prolonged and bitter teachers' strike in 1992 had left the teachers and the board in our district with a great deal of mutual mistrust. During discussions with our staff, we revealed that the Modified Four-day Week proposal would involve 20 more minutes per week of instruction than the present system. Over a semester this would add up to 20 additional hours of instruction, or more than three full school days. We included this in the Modified Four-day Week proposal to deal with the possible argument that it would be just another day off for teachers. We simply felt the extra instructional time would be more saleable to the public. However, the bitterness of the strike had the staff focus on the instructional time issue extensively. The strike had left some staff members feeling unappreciated and bitter, and they didn't like the idea of working extra hours for no remuneration. In addition, the inevitable 5% salary reduction proposed by the Alberta government in its cuts to educational funding caused some staff to immediately dismiss the idea because the proposal would see teachers ultimately working more hours for less money. After a long debate, the staff voted by secret ballot on the proposal. The result was that just less than 80% of the staff endorsed the idea. My personal perception was that many teachers felt that the board would just reject the idea anyway. Regardless, this staff endorsement meant the proposal would go to the board.

Many hours were spent preparing the presentation that our principal was to give to the board of education. For the first time research was solicited, specifically from Jim Weed, the principal of Meadowbrook School in Airdrie (a small city just north of Calgary, Alberta) who had been working on a four-day school week proposal for his school for over a year. Hours were spent combing the research to bring out both the positives about the four-day week and also the concerns. We were pleased with our findings. The four-day school week had been successfully implemented in a number of school districts in the United States, specifically in the states of New Mexico, Colorado, Oregon, and New Hampshire. According to separate studies by Grau and Shaughnessy (1987) and Reinke (1987), the positive benefits of the four-day school week were numerous. First of all there were significant savings (Grau & Shaughnessy, 1987). Reinke (1987) indicates that "transportation and maintenance costs were down 18%, bus gasoline savings from 15% to 23% and there were additional energy savings (electricity, fuel oil) as well as classified employee's salaries (bus drivers, cooks, secretaries, aids, and custodians)" (pp. 3-4). Given the nature of Alberta politics with the emphasis on eliminating the provincial budget deficit through program cuts, we knew that this would be a big selling point with the politicians.

Other positive factors were identified in the research such as more time for staff development, more time for co-curricular and extracurricular activities, improved attendance by staff and students, higher school morale, more parental involvement, often more instructional time, more family time, and lower dropout rates (Reinke, 1987, pp. 4-5). In addition, McCoy (1983) states that there is "no reason to believe that school achievement or quality of education is hindered when school districts switch to a four-day schedule" (p. 35). In

endorsing the four-day week, Blankenship (1984) states, "Teaching 'smart'—that is, making more efficient use of available learning time—is the best way to promote effective learning" (p. 32).

There were also some drawbacks to the four-day week. Reinke (1987) identifies the following drawbacks: school financial savings are downloaded to parents (babysitting, etc.); longer days are difficult for younger children; loss of retention occurs for special-needs students; a number of three-day weeks result from other holidays; the difficulty and expense to revert to a five-day week; and the inconsistency of the four-day week with a strong North American educational reform movement that advocates for even longer school days and a longer school year (pp. 8-9). We felt because we were a junior high, dealing with students aged 12 to 15, that many of these drawbacks, especially child care, would not necessarily apply, and we were enthusiastic about the possible positive benefits of the four-day week.

We also felt that our board was confused, as was everyone in education in Alberta, about the direction of funding, and that they would be more willing to experiment than they had been in the past. This feeling, along with the research, made us cautiously confident that our proposal would be accepted despite the history of changes with this board.

The presentation to the board went well, and late in the afternoon we were notified that the board had passed the motion to permit the pilot project subject to "parental approval." Panic set in with our administrative team. Parent approval was far more elaborate than our original plan. It implied a general meeting and a formal vote, not an item on a Parent Advisory Committee (our school parent group) agenda! The date was January 11, and we were scheduled to begin February 1. Time was of the essence, and neither parents nor students had been contacted.

A Huge Error?

Those who teach in a secondary school will understand—it's tough to get students to take home information. We had to get the students to take a letter home to their parents notifying them of a public meeting to discuss the four-day week due to the board motion. We had timelines; therefore, the students had to be informed of the change so that they would make sure that the letters got home. If we didn't tell them, we felt that the letters would not get home. Either way, the process was wrong from the parent point of view. Don't tell the students, and the information does not get home—a mistake. Tell the students so that the information gets home, and the parents are angry at being the last ones informed—a mistake. Mail the letters and we've spent $250.00, and the information may not get home in time given our deadlines—also a mistake. Well, we chose to explain the changes to the students. Our administrative team divided up the classes in the school, went over the positive and negative aspects of the change with the students, and took a student vote. The result was that 82% of the student population was in favor of the Modified Four-day Week. We then gave them the letter outlining the project and inviting the parents to a general meeting to discuss the project. Now we had to sit back and wait.

The letter that we sent home was hastily prepared and, to be honest, not very good. The next day our monthly newsletter came out, and we tried to address the problem of the initial letter in a more diplomatic manner, but we knew it was too late. Conversations with our Parent Advisory Committee (PAC) President revealed that she was squarely against the proposal. She had not been informed, she not did view the proposal in the spirit of an educational partnership, the process was wrong, and she proceeded to oppose us in the local newspaper. With five days to the public meeting, we were going off the rails.

The Public Meeting

The unsettling thing going into the public meeting was that we had received only one phone call and virtually no response from the public regarding the pilot project. The responses that I solicited from friends were varied. I was accused of selling the kids a bill of goods and then encouraging them to go home and work on their parents, and I was praised for trying to be innovative. I did get a chance to practice both my debating skills with the research and trying to remain calm during verbal assaults. The second part was a lot tougher.

Preparation for the meeting was taking time away from the problems in the school and the classes that I teach, and the pressure was taking away from my sleep. Basketball coaching was not helping the problem. Fullan was right. Change is tough. Somehow, reading and writing about change was a lot easier than implementing and surviving it.

Before the meeting we knew a number of things would be difficult. First of all, the PAC President would probably breach the impartiality of the chair and enter debate against us. This happened. Also, at the last minute we became aware of a poster campaign against the program that was circulating around town bemoaning the fact that the students would miss intramural time, lunch time, and noon-hour club time if the proposal went through. We could clearly see where some individuals' priorities in education were going to be.

At the meeting our principal spoke first regarding the new program in philosophical terms, and I spoke about the actual timetable and calendar and also addressed the difficulties of changes. When we were both finished, to absolutely no applause, we were ready to field questions. The first question was a dandy: "If these three extra days that you are going to be working as a result of this program change are so important, does that mean if this proposal doesn't go through, that you will take a day off Easter Break, Christmas Break, and summer vacation to make them up?" Oh boy, I could feel my temperature rise, my stomach churn, and total despair consume me. I wrote a note on my outline: "It's dead." The audience seemed to think it was a relevant question, however, and it summoned some courage in two women in the front row to continually roll their eyes and provide negative commentary on any of our responses. This was personal, and we were alone. Much of the rest of the evening was a blur. Sometimes it seemed like attack after attack, cheap shot followed by cheap shot. By 8:45 I was exhausted and depressed. I had come to school at seven that morning for a basketball practice, worked through the day, including detention room at noon, had meetings after school, gone home for 45

minutes for supper and to change my shirt, and now we were being attacked for being lazy teachers—9:00 to 3:30 and wanting an extra day off every two weeks. We refused to be baited by the attacks, and we stuck to answering the questions. I wanted to scream that an eight-hour day for me ended about five hours ago, but some of these people had not come to listen.

Then it happened—a member of the audience spoke in favor of the proposal. She said that she thought it was worth a try. She appreciated the time and effort put into this proposal, and that the change in work week might lead to better academic performance and significant savings. Besides, she reasoned, five months was not such a long time to try something, and most of the months had only one extra day off because of all the Friday holidays anyhow. She sat down to applause. You could feel the mood in the room start to change. There was a recognition by some members of the audience that we had taken a lot of abuse, and that we had anticipated this abuse coming into the meeting. Despite this unpleasant pressure, we had pressed on. I think a number of the audience began to appreciate that it would have been much easier for us to do nothing and spare ourselves this meeting. Those people were taking the proposal very seriously.

After the meeting, I accepted some compliments on both my courage and the idea. To some we had become martyrs in the process because some of the attacks had been too harsh. The final straw for many occurred when a parent suggested that the ballot box be carefully scrutinized, implying that we couldn't be trusted to count the ballots fairly. One parent immediately stood up, apologized for the slight, and added, "We trust them with our kids. I am sure they can count the ballots!"

I was just glad it was over. Our administrative team met and vowed that tomorrow we would forget this meeting and concentrate on the school.

The Next Week

The next day we were out in force with the students and in the school trying to atone for a week of neglect. We were determined to focus on the school; however, it was not entirely possible. Our PAC President insisted on a great deal of control over the material that was to be sent out with the ballots to the parents. She had gone to central office expressing her concern with the process. Our principal was called over to central office and learned that they had some second thoughts over the process. The letter that was to be sent with the ballot had to be changed to a neutral stance, and the ballot would now have three options: yes, no, and delay. There would also be room for comments on the ballot. When he returned to the school, the pressure of the situation had taken its toll. He had taken this project through central office, through the board, through our staff, through the students, and now he suggested scrapping the ballot and shelving the project. He had shown tremendous courage and creativity to construct a workable plan within a system with a large busing component and shared instructional facilities. He was demoralized because he felt that central office feared the project would fail, and he was uncomfortable with the fact that some groups, notably parents and bus drivers (understandably with the future possibility of losing a day's pay), would try to sabotage the pilot and not allow it to work.

At this point, I reminded him of all that we had gone through to this point, especially the public forum, that we had indicated that there would be a ballot, and that parents would have final say on the project. Our credibility had been attacked, and if we bailed out now by shelving the project we would lack credibility in the future. We finally agreed that, win or lose, the vote must take place. In addition, we would invite other groups to help us count the ballots because we did not want to be accused of tampering with the results if we counted them in private and we won. We wanted nothing to do with those ballots.

The PAC President and Assistant Superintendent agreed to help tally the results, and one more final spin was put on the vote at the insistence of the PAC President: the result was only valid if over 50% of the parents responded. Anything less than 50% would mean defeat of the proposal regardless of the count. This increased the tension with the PAC President, and we were all stressed.

The Ballot Count

I had a basketball game after school so I excused myself from the ballot count. In truth, I wasn't sure I cared anymore. I secretly vowed that I would not go through this again. I had relearned the lesson of the first-year teacher: theory and practice are sometimes miles apart. Change is a lot tougher than Fullan indicated, and he never said it was easy.

The final ballot had a 58% response rate with 58% of the respondents voting "Yes," 36% "No," and 6% "Delay." A local television station had set up at the school to give both radio and television coverage to the announcement of the vote. Our principal thanked those parents who supported us, and vowed do to all we could to work together with parents who were opposed to make the project work. Our PAC President predictably criticized us for the process saying parents weren't properly involved or informed. Personally, I was surprised by the outcome. I had prepared myself for a defeat psychologically so I wouldn't feel as much bitterness if the proposal were rejected. I then experienced a series of different emotions. First I was euphoric about the victory of our proposal; however, suddenly I became awestruck about what we had accomplished amid such negative and vocal criticism. Then panic set in as I realized that we were to begin going under the microscope for the next five months, especially with over one third of the parents opposed to the plan.

"Andy Warhol's 15 Minutes"

The next school day was chaotic. Two television crews wanted to come into the school and report on the project. Numerous radio stations from all over Canada requested interviews and quotations. School systems and schools phoned requesting information regarding the pilot project. It was a day of images I'll never forget: dozens of hyper adolescent students followed an overwhelmed television cameraman around the school at noon hour as I was rushed off to do an interview with a "Talk Radio" station in Winnipeg. I was physically and mentally spent, and I wondered when and if school would ever be the same.

The Aftermath

We have been on the new program for two months, and so far there have been a few surprises. The first surprise was the initial burst of interest in the project. Because we had been so preoccupied with the pilot, we didn't focus as much on the events in the provincial scene. In conversations with other educators, I didn't realize how prevalent the feeling of despair was in the ranks of teachers and administrators throughout the province of Alberta because of the Klein government's direction in education. Our project had served as a diversion from the gloom and doom in the Alberta educational scene.

The second minor surprise was the criticism that our project received from some educators in the area, specifically teachers at the elementary level. Some teachers saw child care as the most important issue in the project, and asked what working parents would do if the program was expanded to the elementary level. I was a little puzzled. If the project were to be expanded, it would be because it was effective and efficient. Supplying babysitting should not be the priority, and I felt teachers especially would recognize this. With the Alberta government pledging to cut nearly 15% from the education budget for the 1994-1995 school year, it is obvious that fundamental changes are going to take place in the schools. For the County of Lacombe the changes will be even greater. The county has benefited tremendously financially from the nature of local funding of education in Alberta because of the large corporate industrial sector located in the eastern section of the county that has pumped millions of dollars of tax money annually into the educational budget of the county. With the Alberta government decision to take over the funding of education from the local governments and then allocate the funds across the province on a per-pupil basis, the corporate funding for the County of Lacombe will be lost. Residents of the county may be faced with the unique experience of seeing their property taxes rise significantly at the same time as their educational budget significantly shrinks. Programs will be lost. Students will suffer. Education will suffer. I would hope that parents would be willing to bear the burden of arranging periodic babysitting if a proposal can help save educational programs for their children or make class sizes more manageable. I would hope that teachers would support any project that may be beneficial for education.

Our first PAC meeting was a surprise also. Despite the split opinions in our PAC regarding the idea, they were extremely helpful in putting together a monitoring process to evaluate whether the project should proceed past June. The monitoring process consists of two parts. The first is a questionnaire that will check the perceptions of parents, students, and staff twice throughout the pilot period. The questionnaire will ask not only if the stakeholders like or dislike the project, but will also solicit input regarding the positives and negatives, and it will ask whether the project should continue. The second part of the monitoring process will deal with the collection of statistical data. The PAC asked that information be gathered on absenteeism rates for staff and students, utility cost comparisons, incidence of discipline problems, a school vandalism report, student achievement comparisons, a police report, and a report from the town recreation department. After the second questionnaire is issued in early June, all the data will be tabulated and forwarded to the Board of Educa-

tion. A report from our principal will accompany these data with a recommendation for the Board. The Superintendent and the Board will make the final decision whether the project will be extended. There will be no more public meetings! I was extremely pleased that even though some of parents on the Parent Advisory Committee did not support the project, they were willing to give it a chance, and they gave excellent input into the monitoring process.

Overall, I think that the project has been successful thus far. Discipline problems related to noon hours have decreased substantially. My perception is that teacher stress has decreased. The tumultuous pre-Easter period for which large junior high schools are famous was less hectic. I believe that students and staff have adapted well to the longer instructional day. Has the change had a positive effect on the teaching-learning process? I honestly don't know. Time will tell.

Implications for Educators

This process has provided us with a great deal of insight into the dynamics of educational change, and I would make the following recommendations and observations to those planning major changes.

1. Principals and other school-based administrators have a vital role to play in the implementation of change. They must take a leadership role and be willing to defend their ideas and accept the negative criticism that will be associated with major change. They must be prepared to accept the emotional baggage associated with major change as the change meets resistance from the stakeholders. They must realize that, unfortunately, the easy thing is to do nothing; however, good educational leadership in today's society does involve taking both risks and the responsibility for those risks.
2. School and school district climate is vital to the change process. This pilot probably would have been rejected in previous years. The cost cutting measures of the Alberta government, the uncertainty of the future of local school boards and central offices, and the high level of anxiety in society regarding the economy have made stakeholders more receptive to change. The political realities of each school and each school district are different, and the agents of change must understand the consequences of those realities.
3. As is indicated by numerous authors who write about change, change is a process, not an event. Effective change takes time, energy, and resources. We have now instituted the pilot project of the modified four-day week. The work has just started. It must be monitored closely, and all stakeholders must have their say on whether the program should be cancelled, continued, or expanded. A criticism that can be made about this project is that the timelines did not facilitate the involvement of all the stakeholders properly, especially the parents. This may be true; however, a look at the Meadowbrook project in Airdrie, Alberta, indicates that on occasion we in education can get paralysis from analysis. That particular project had over a year and a half lead time with extensive research and consultation with the stakeholders. It was delayed consistently pending final approval from the Board. It finally passed—less than a week after ours. Coincidence? I think

not, and neither do some individuals involved in the project at Meadowbrook.

4. Make sure to have central office staff and the local board clearly on-side. The vision that you have of the change must also be their vision. If these stakeholders have second thoughts when the heat is on in the middle of the process, this potentially can be devastating to the project. In fairness, the second thoughts we perceived in our central office after our public meeting were neither malicious nor of the nature of bailing out; it's just that we got advice when what we really needed was encouragement.
 Overall, although our board and central office did not take an active public role in the process, they did have a sincere desire to see the proposal work. They also exhibited courage in risking public backlash by allowing the project to proceed when they easily could have destroyed it in its infancy.
5. Make sure change agents have a clear grasp of the big picture. The process can get extremely emotional and personal, and kneejerk reactions to the pressure are understandable. As an administrative team, you need to support each other and be sympathetic listeners when emotions are frayed; however, you must maintain a focus of your original goals and realize that time will inevitably pass and so will emotions.
6. Because change takes a great deal of time, energy, and support, proposed changes must be carefully weighed to consider not only the desired effect of the program, but the effect that the implementation will have on existing programs. Innovation is not simply an addition; it changes the structure of the programs in existence because those programs may have to sacrifice the time and energy that they previously received in order to implement change. Our change has affected the entire nature of the school right down to administrivia such as supervision schedules, where students are dropped off by buses, and the way announcements are made. Also, the time that was taken to prepare for the change by administration postponed other duties. Therefore, proposed changes must be priorities, not frills, or the quality of education offered to the student could be sacrificed.

There is little question that we in education, especially in Alberta, are living in difficult times—times of change. These are not just the economic changes and the impact they will bring, but also changes in parental roles and governance. What we achieved in this project is not a blueprint for change; it may be merely the "science of muddling through" (Lindblom, 1959). However, unless we are willing to take chances and initiate change, we will be forced to accept changes and will relinquish to others control of our destiny. Our students will have to be creative, resourceful, and flexible to succeed in the future. The times dictate that as educators we will have to be creative, flexible, and resourceful to ensure quality education in the future. We can no longer justify systems in education based on the reason that "That's the way that it's always been." Our modified school week may prove to be no more than a five-month experiment, but it has enabled us to explore and discover things that we would not have discovered had we done nothing. I also have no doubt that some of these discoveries will benefit the students in our school. To me this makes the project successful whether it continues or not. As educators, we also can no longer take the easy way out and avoid conflict. In our current educational climate, taking

the easy way out may be the most dangerous route we can take with the futures of children.

References

Blankenship, T. (1984). Update: These schools swear by the four-day week because students work harder and face fewer distractions. *American School Board Journal, 171,* 32-33.

Fullan, M. (1982). *The meaning of educational change.* Toronto: OISE Press.

Grau, E., & Shaughnessy, M. (1987). *The four-day school week: An investigation and analysis.* ERIC Document Reproduction Service No. ED 317 337.

Hord, S. (1987). *Evaluating educational innovation.* New York: Croom Helm.

Lindblom, C.E. (1959). The science of muddling through. *Public Administration Review, 19,* 79-88.

McCoy, J. (1983). *A summary report on the four-day school week in New Mexico.* Santa Fe, NM: New Mexico State Department of Education, Evaluation, Testing, and Data Management.

Reinke, J. (1987). *More with four: A look at the four-day week in Oregon's small schools.* Salem, OR: Oregon Department of Education.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 283-296*

David Watt

and

Hetty Roessingh
University of Calgary

ESL Dropout: The Myth of Educational Equity

Research into the educational equity afforded students in Canadian schools has often masked the case of ESL high school students in larger statistical samples. This article reports the findings of a unique tracking study aimed directly at ESL high school students. The study tracked the educational progress of 232 ESL high school students and found an alarming blended dropout rate of 74%. It also calculated specific dropout rates according to the one consistently predictive variable that emerged from the study: English language ability at entry to high school. Apparent differences in the patterns of dropout are further characterized under the labels fall-out, push-out, and dropout. The article concludes with some recommendations to address the myth of educational equity for ESL students in high school.

La recherche dans l'équité éducative des élèves inscrit(e)s dans les écoles canadiennes a souvent masqué, dans les échantillons statistiques, le cas des élèves du niveau secondaire qui apprennent l'anglais comme langue seconde (ALS). Dans cet article, on présente les conclusions d'une recherche unique qui cherchait à dépister le tôt de décrochage des élèves ALS du niveau secondaire. La recherche a suivi le progrès scolaire de 232 élèves du secondaire dans un programme ALS et révèle un tôt alarmant de décrochage scolaire combiné qui s'élève jusqu'à 74%. L'habileté langagière en anglais dès l'entrée au secondaire est une des variables ressortie de la recherche-même qui pouvait prédire de façon constante le tôt de décrochage. En utilisant cette variable prophétique on a pu calculer le tôt de décrochage spécifique des élèves ALS. Des difficultés apparentes dans les modèles de décrochage sont d'autant plus caractérisées sous les étiquettes: "fall-out" ceux et celles qui quittaient les rangs de façon volontaire; "push-out" ceux et celles qui se faisaient sortir du programme; et "drop-outs" ceux et celles qui se retiraient ou décrochaient du programme. On conclut l'article en présentant des recommandations qui adressent les mythes de l'équité éducative pour les élèves ALS des écoles secondaires canadiennes.

Introduction

Immigration and the educational success of school-aged immigrants are essential elements of the Canadian ethos. Immigration has been on the rise since 1986, and all indications are that it will continue to increase at a rate similar to that of the great immigration waves of the 1900s. The interests of immigrants and Canadian nationhood converge: although immigrants may choose to come to Canada, Canada is also economically dependent on the successful integra-

David Watt is an assistant professor in Curriculum and Instruction, where he teaches undergraduate and graduate courses in the area of ESL. His current research interests include educational equity, language minority student education, and second language literacy.
Hetty Roessingh is a doctoral student in Curriculum and Instruction. She is currently working on curriculum revision for high school ESL programs. Her research interests include the development of English for academic purposes and content area language teaching.

tion of immigrants (Statistics Canada, 1993). New Canadians stabilize the population base and ensure sustainable economic growth (Employment and Immigration Canada, 1990). Essential to this growth is a well-schooled, socially cohesive, literate labor force. The debilitating effects of limited literacy, educational dropout, and the resulting social marginalization fray the social fabric.

A shift in immigration trends over the past 15 years has given rise to a number of educationally related policy and programming initiatives. The 1971 multicultural policy recognized the essential role of language, culture, and ethnicity in the Canadian identity. At the same time, English as a Second Language programming became an essential part of the institutionalized responsibility of education. At the local level in Alberta, the Calgary Board of Education provided ESL support to 3,338 students in 1993, one third of whom were either junior high or senior high school age. How do these students fare in our school system? Research on immigrant children's successful integration into the education system has focused largely on academic achievement. Canadian studies (Cummins, 1986; Early, 1993), as well as American studies (Collier, 1987, 1989; Collier & Thomas, 1989), have investigated the level of language proficiency required for academic success, the length of time required for immigrant students to become academically competitive with their English-speaking age peers, and aspects of achieving academic success. However, academic success is the combination of both academic achievement and year-to-year educational progress. Although achievement can be measured by testing, progress must be measured by tracking. Progress studies that report statistics on early school-leavers tend not to focus on the progress of high school ESL students. Radwanski's (1987) study of high school dropout in Ontario has noted a consistent increase in ESL dropout reaching the level of 53% in 1986. More recently, Alberta Education's study (1992) reported a 61% dropout for ESL secondary students in Alberta.

Although these studies give reason for alarm, we felt that a refinement of these findings was necessary for them to be useful for the purposes of policy development and program design. The issue that lies behind these studies is that of educational equity. A growing number of studies (Grey, 1991; Minicucci, 1992; Sinclair, 1992; Spener, 1988) indicate that language minority students do not experience academic success in school, but rather suffer from frustration and marginalization, finally culminating in dropout. The aim of this study is to provide some insights into the educational opportunities that exist for ESL students in the present system in the province of Alberta.

Purpose

The purpose of this study was to track the progress of ESL high school students from their first ESL class to their point of exit from school in such a way that we could:

- specify points of exit, degrees of risk, and profiles of school leavers;
- identify significant factors among those who successfully completed a high school diploma;
- make some suggestions for change in policy, administration, and programming, based on the findings.

Background

The definition of dropout varies widely in the current literature, though this has not always been the case. Until only recently *dropout* was broadly defined as "anyone who did not complete the requirements for high school graduation." Education Ministries across Canada as well as other stakeholders uniformly accepted this definition, and thus there was consistency in the way dropout was calculated and reported. A generally accepted figure for dropout nation-wide is 30%-35% (Human Resources and Labour Canada, 1993).

An examination of recent local, provincial, and national studies reveals a multiplicity of definitions and methodologies for calculating dropout (Alberta Education, 1992; Calgary Board of Education, 1991; Human Resources and Labour Canada, 1993). Often these are related to the sources of data available for such studies and include high school attendance records, graduation rate data, and family allowance data. Accurate reporting and hence understanding of dropout is beset by recent methodological innovations. New methods of calculating dropout easily confound a naive public. The *annual (per grade) dropout* rate calculates the percentage of students across a particular age or grade range who drop out in any given school year. This figure tends to be low, averaging approximately 8% over the past decade in Alberta (Alberta Education, 1992). *Longitudinal dropout* takes into account students entering grade 9 who drop out before completing high school: thus a cumulative calculation. This figure is reported as 34% in Alberta (Alberta Education, 1992). The terminology *confirmed* and *apparent* dropout are used to refine the construct of dropout at the jurisdiction level (Calgary Board of Education, 1993). Confirmed dropout includes those students who participated in formal school withdrawal procedures (26% system-wide for grades 9-12: 1991-1992). Apparent dropout includes the larger number of students who do not make their intentions or plans known and who cannot be traced in the system (41% system-wide for grades 9-12: 1991-1992). Administrative personnel in the Calgary Board refer to a *blended* dropout rate that would reflect an average of the confirmed (lower figure) and apparent (higher figure) dropout. This would result in a dropout figure of 33%.

Further, the introduction of new terminology to account for students who disappear over the summer, students who drop out but then come back, or students who register as adults in postsecondary schooling institutions all contribute to the manufacture of "dirty data" (LeCompte, 1987), and the consequent difficulties in comparing study results. But to do the math is to know that in essence little has changed over the past decade: the dropout figure (i.e., the number of grade 9 students who do not fulfill graduation requirements) remains at 33%.

Little is known about ESL dropout in the context of these figures. Once again, the problem of definitions and methodology confounds the results, making it difficult to compare studies. Alberta Education (1992) proffers a blended dropout figure of 61% for students between grades 8 and 12, based on 1987 funding figures. Radwanski (1987), in his Ontario study, suggests a dropout rate of 53% for high school ESL students (grades 9 to 12).

For the purposes of this study, *dropout is defined to include all grade 10 -12 ESL registrants who withdrew from school without having fulfilled the requirements for*

graduation from high school. This simplified definition of dropout is largely consistent with the definition offered by the Calgary Board (1993) in the *Student Withdrawal Report, 1991-1992 School Year*, and allows us to focus on the real problem of early school leaving. As Radwanski (1987) succinctly put it, "divorce is divorce." Whether one wants to play the game again at a later time is another issue.

ESL is defined to include all students who received ESL support while registered at the high school at which the study was conducted. This included students who fell within the guidelines established by Alberta Education as outlined in Part 2, Section 9 of the *School Grants Manual* (Alberta Education, 1992), as well as students who did not meet Department criteria for funding: that is, they were beyond the three-year funding limit or were Canadian-born but nonetheless were offered support on the basis of demonstrated and obvious need.

Method

The study was conducted in a large, comprehensive urban high school in Alberta. Of the 1,500 students registered at the school in 1992-1993, 40% spoke a language other than English as their first language; however, only about 8% of the school population received ESL support during the years of the study (1988-1993). The most-often spoken languages other than English included Vietnamese (12%), Chinese (10%), Arabic (7%), Spanish (3%), and Punjabi (2%).

Between 1988 and 1993, 388 students had been involved in the ESL program. All these students were tracked from their first day of attendance at the school until they left. The educational status of 232 of these students (those enrolled between 1989 and 1991) can be completely accounted for. The remaining 156 students are either still attending the school (i.e., 107 students from the intake groups of 1992 and 1993) or constituted the intake group of 1988 (49 students). The latter group could not be tracked for annual progress because of incomplete records in 1988. However, we were able to determine their high school graduation status in 1992 or 1993. This study was concerned with both achievement *and* progress, and the 1988 class cohort contributed significantly to our understanding of the phenomenon of achievement.

The students are a diverse group: from sophisticated, big-city Beirut to the camps of Bangkok; from talented academicians to illiterate refugees. They have arrived from 43 different countries and speak 29 different first languages. Certain students arrived as younger children and had already spent some years in their quest to learn English well enough to compete academically with their Canadian age peers. These students tended to be beginners on arrival. Those who were older on arrival represented enormous diversity in terms of first language proficiency, educational background, and proficiency in English.

The ESL program at the school serves the needs of approximately 115 students each year. It is based on the philosophy of proportional integration; that is, it aims to move students from beginner level (those who may need full-time ESL support) to advanced level (those who are integrated 75% of the time into the mainstream) within the three-year time allotment granted by Alberta Education. The express goal of the program is to prepare ESL students from the perspective of language proficiency and content knowledge for the demands of the mainstream academic program and thence toward high school graduation. The structure and sequencing of courses can be flexible to suit the

needs of the students. That is, students work through the program at their own pace, with the possibility of skipping courses along the way if they are seen to be ready and competent to accept the challenge of more difficult course work. Generally, however, from beginner to advanced level takes a full three years. In fact, few beginners are able to make the sort of progress required to move into the mainstream within three years, as our findings show.

Procedures

All ESL students who attended the school between 1988 and 1993 were tracked from the time they were first registered at the school until they withdrew. ESL funding lists and teachers' class lists from the years 1988-1993 readily facilitated the completion of this initial task. A summary of the numbers enrolled in the program appears in Table 1. The new intake figure as well as the total number of students in the program each year is displayed. This became an important piece of information.

Various factors thought to be predictive of academic success were identified and cross-tabulated to determine which of these would be most useful as a basis for tracking the ESL students' progress in school. Among these variables were gender, age on arrival, first language (L1) ability at entry, L1 educational status, country of origin, home language(s), number of languages spoken, educational gaps, immigration status, and English language proficiency at entry. Some of the variables had to be dropped because records were insufficiently complete for the group as a whole. In the end, three variables were discarded for logistical reasons. The number of languages spoken, the educational gap, and the immigration status were discarded because of inconsistent reporting or unreliability. We treated another variable, L1 educational status, as only moderately reliable. It was calculated from graded math tests given in L1 at entry to the school board and operated on the assumption that an ability in mathematics is indicative of some form of educational experience and status. From the remaining variables, English language proficiency on entry into the school emerged as the most powerful predictor of success, and thus it was used as the basis for tracking all students who received ESL support.

A class cohort profile based on language proficiency (beginner, intermediate, or advanced) was developed for each intake group for the years 1989-1993 (see Figure 1). This information was gleaned from student files and teachers' class lists. For a variety of reasons, including inaccessibility of ar-

Table 1

	New Intake	*Total Cohort*
1988	49	105
1989	55	92
1990	85	123
1991	92	137
1992	58	115
1993	49	85
Total	388	

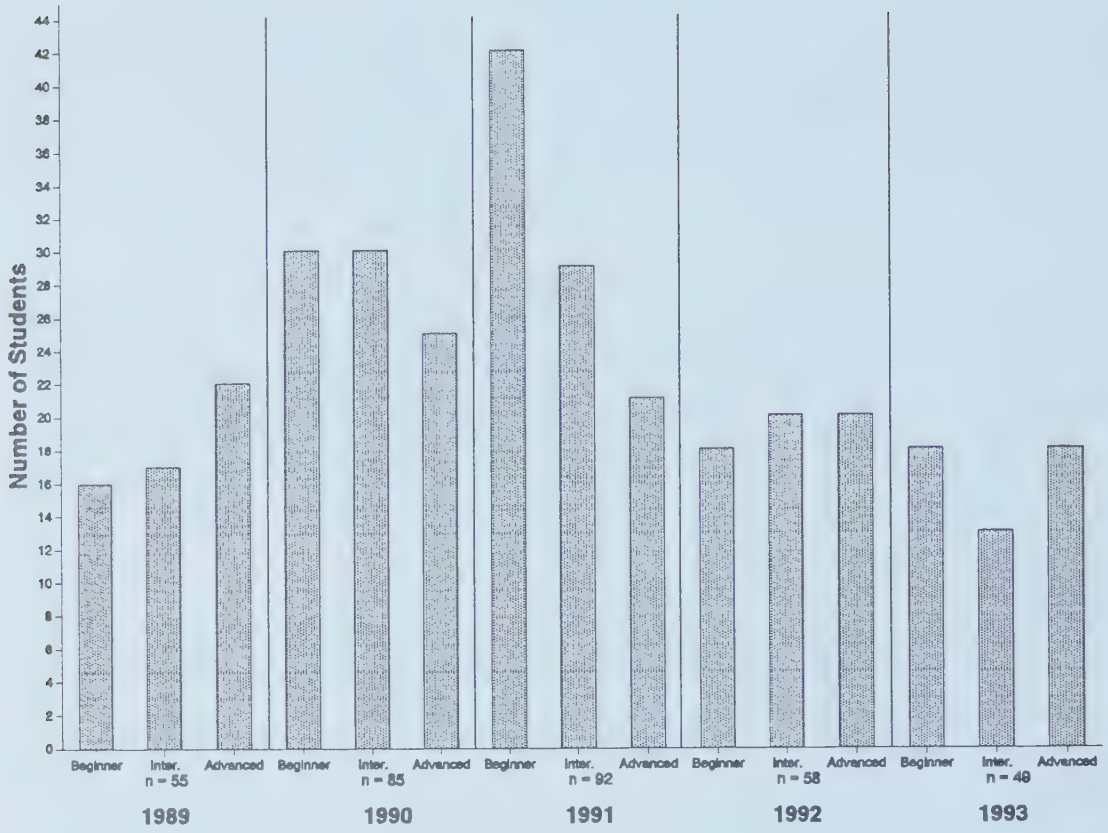


Figure 1. Intake groups: 1989, 1990, 1991, 1992, 1993.

chival material and computer dumping of class lists, we were only able to access this information as far back as 1989.

All the students who entered the high school system between 1989 and 1991 could be accounted for, and so this group of 232 students became the focus of the next phase of the study. Each of these students was tracked every September to note educational progress until we could no longer find him or her in the system. This was accomplished by accessing the students' computerized cumulative record through the use of the students' system identification number (students also have an Alberta Education identification number, and this number is important for tracking academic achievement among graduates). Computerized student records were then matched with individual student files in order to guarantee that we could account for every subject in our study rather than discarding individual cases for reasons of technical difficulty.

The changing shape of the class cohort profiles based on language proficiency (Figure 2) was noted for the 232 students (1989-1991).

The educational progress of the cohort is complicated by students who take varying amounts of time to complete a level. The actual tracking progression was described as follows. In September 1989 there were 16 *beginner* students in the intake group (students are not grouped by grade or age, but rather by language proficiency). By September 1990:

- eight had disappeared from the system,
- seven had advanced to intermediate level, and
- one student had moved forward into advanced level.

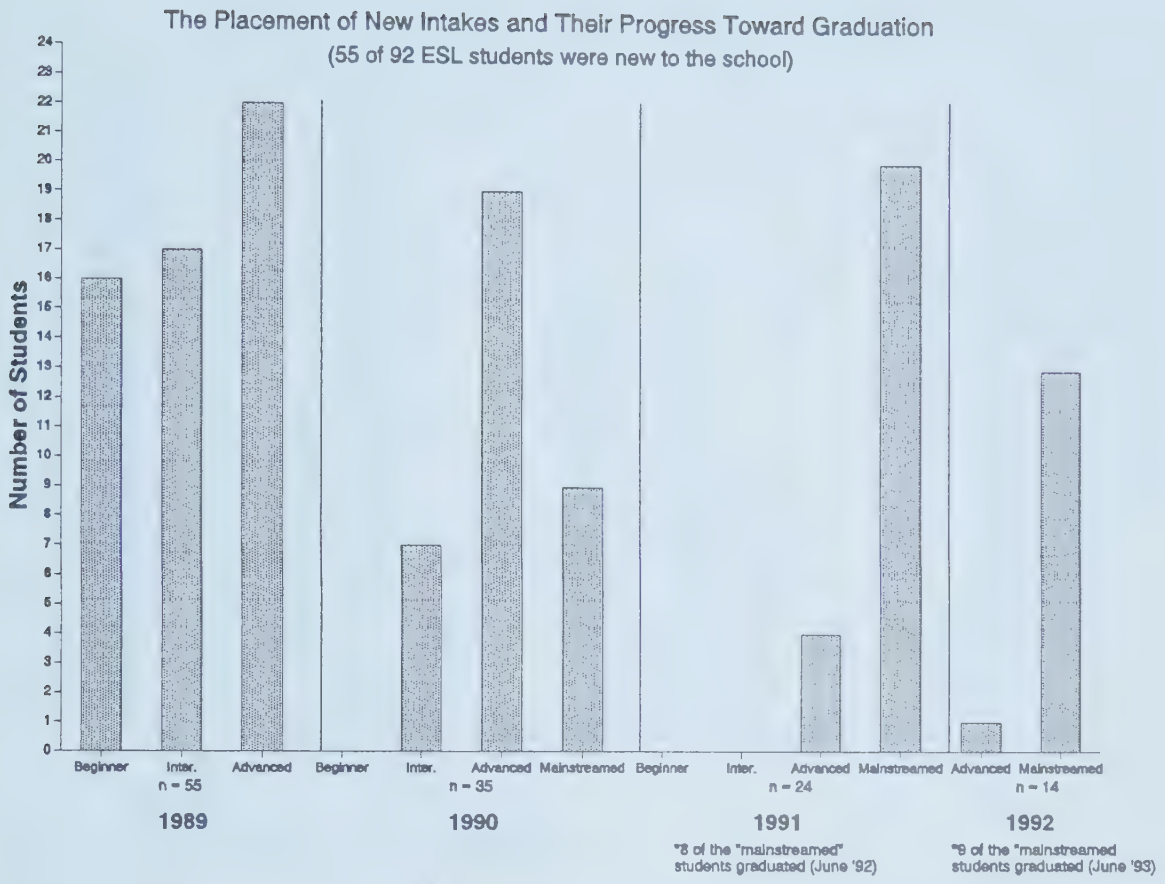


Figure 2. Class of '89.

Of these eight who remained, September 1991 revealed the following:

- four had disappeared from the system,
- three moved forward into advanced level, and
- one student had moved forward into the mainstream.

Looking once again in September 1992 we found:

- one had disappeared from the system,
- one remained in advanced level ESL, and
- two were in the mainstream.

Three of the original group of 16 now remained in the school. In September 1993, we found that:

- one student had dropped out of school;
- one student was still at the school but was far from graduating;
- *one student graduated in June 1993.*

We repeated this procedure for the intermediate and advanced groups of 1989 and again for all the students in the class cohorts of 1990 and 1991. All 232 students were accounted for.

Figures 3 and 4 display the progression of the class cohorts of 1990 and 1991.

Next we wanted to focus our attention on those students who had graduated from high school. We turned to the graduation lists of June 1992 and June 1993 to identify those ESL students who had successfully fulfilled Alberta high school graduation requirements. We then accessed the students' guidance files to see if we could identify the constellation of factors predictive of academic success among ESL students. There were 21 ESL students in the graduating

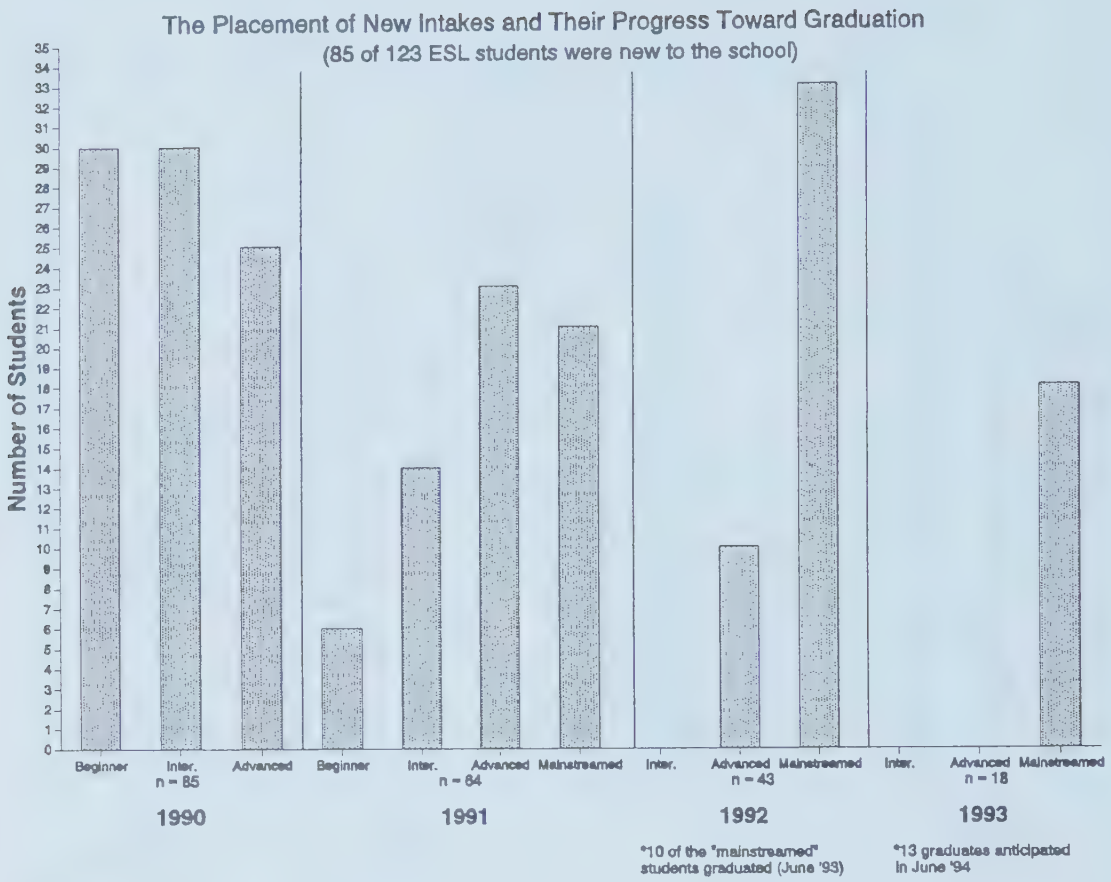


Figure 3. Class of '90.

class of 1992 and 23 in the class of 1993: a total of 44 graduates. Their complete student guidance files were studied in depth, where from rich background information the following variables were extracted.

- country of origin,
- first language proficiency,
- proficiency in English,
- educational achievement (L1 math ability, years of schooling),
- date and age on arrival,
- initial grade placement,
- cumulative student record,
- Departmental exam results (from Alberta Education),
- guidance counsellors' interview notes.

Findings

We discovered that the profile of successful diploma recipients was as mixed and confusing as Early's (1993) small study had suggested. Early in the study, however, the variable "English language proficiency at entry" emerged as the most likely predictor of success, hence the decision to track the class cohorts on this variable. In addition, we noted the following:

- 86% of the graduates were at the intermediate or advanced level on entry to high school,
- they took an average of 4.5 years to complete the three-year high school program,
- more than 50% were over the age of 20 at the time of graduation,

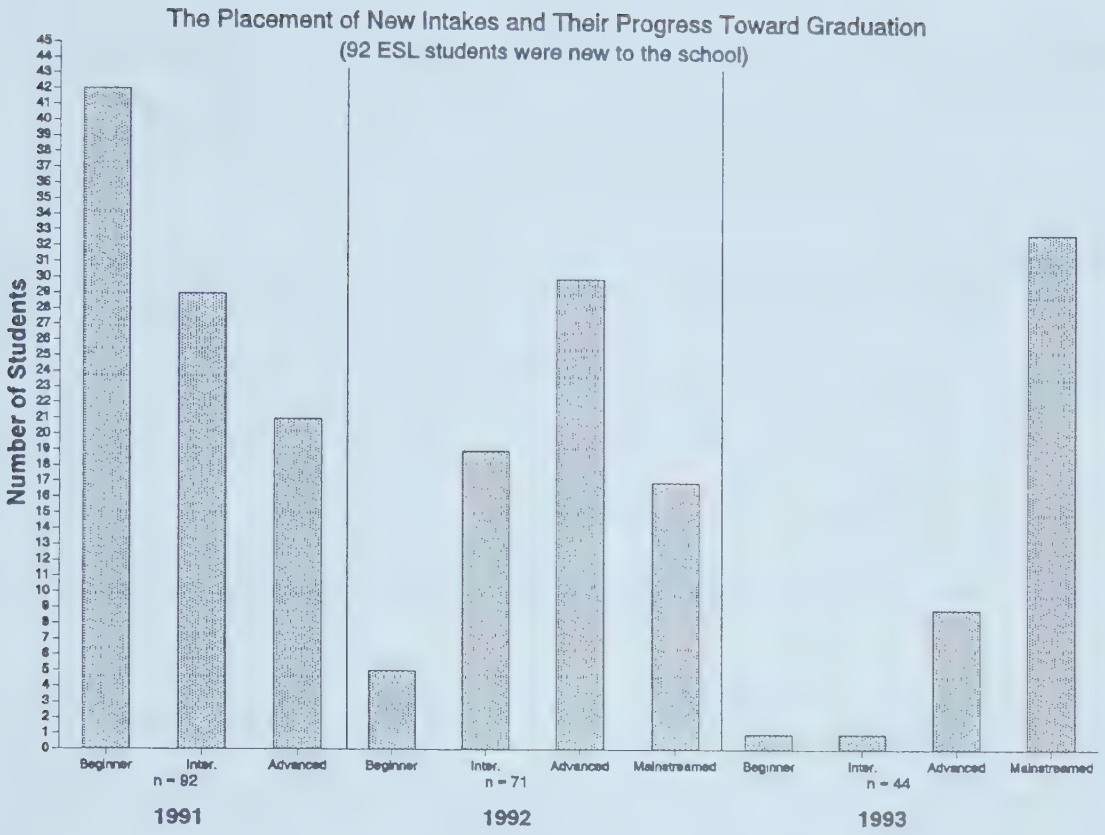


Figure 4. Class of '91.

- the number of years of support for the successful ESL graduates varied according to initial language ability. Beginners averaged four years of ESL support, intermediates required 2 years, and advanced students required only a year of direct support,
- success was directly related either to the number of years of support or the student's language and educational ability on entry into high school,
- about 90% of the graduates received a General Diploma, not Advanced. Among all graduates in the system, approximately 42% received a General Diploma, and 55% an Advanced Diploma.

Most striking was the finding that the overall dropout rate for the 232 ESL students was 74%, a figure substantially higher than that emanating from Alberta Education's study (1992) of ESL students. Further, the ESL students appeared to drop out of high school at twice to three times the rate of their Canadian counterparts in Alberta, earlier cited at 34%.

More striking still was the fact the ESL dropout rate itself varied dramatically according to the student's English language proficiency at entry to high school. Students who started high school with a beginner level of English proficiency suffered a 95.5% dropout rate. Those few exceptional students who started high school with an advanced ability in English had only a 50% chance of completing the high school diploma. Table 2 displays the numbers of beginners, intermediates, and advanced students from the class cohorts of 1989, 1990, and 1991 who graduated. From these figures we were able to develop a

Table 2
Developing a Formula for Calculating Academic Success for High School ESL Students

	Beginners	*Intermediates*	*Advanced*	*Total*
1989	16 6.25%, 1 graduated	17 29%, 5 graduated	22 50%, 11 graduated	55
1990	30 0%, 0 graduated	30 30%, 9 graduated	25 56%, 14 graduated	85
1991	42 7.1%, 3 graduated	29 31%, 9 graduated	21 42%, 9 graduated	92
Total	88 4 graduated	76 23 graduated	68 34 graduated	232
Total Percentages	4.5% will graduate	30% will graduate	50% will graduate	26%

formula for calculating academic success for the ESL students in our study represented in Table 2.

To summarize, we found that 95.5% of the beginner level students dropped out prior to attaining a high school diploma; 70% of the students who started as intermediates also dropped out, and 50% of the ESL students who entered at an advanced level did not complete high school.

Perhaps the bleakest finding was discovered in tracking the ESL students as they progressed along the line of proportional integration from beginner to intermediate to advanced and into the mainstream. Here we discovered that 50% of the beginners would exit the educational system permanently sometime in the first year of registration of high school.

Our study also indicated a trend toward more beginners in each annual intake group and that some of these beginners now needed more time at the beginner level for reasons of limited first language literacy or previous lack of educational opportunity. We noted this trend beginning in the class cohort of 1990 and onward.

A recent reduction in ESL program support at the municipal level has led us to reexamine our data and predict even fewer high school ESL students among the graduating classes of the future: 94% of the graduates of 1992 and 1993 in our study either required more than three years of ESL support or were over the age of 19 in their grade 12 year. Table 3 summarizes the impact of a three-year funding cap or an age cap of 19 years of age for high school attendance on the academically successful ESL students.

Note that it is the younger-arriving students who require ESL support beyond the three-year cap. This finding is consistent with those of Collier (1987) and Olshtain (1990) who posited that first language proficiency is one of the factors predictive of academic success in the second language. Lack of cognitive academic language proficiency (CALP) in L1 means that the acquisition of CALP in the second language will take time—if indeed the student ever acquires native-like age-appropriate proficiency in L2 (Cummins, 1986). Older-arriving students who were academically competent in their first language and

Table 3
The Impact of Imposing Funding and Age Caps on Academically Successful ESL Students (1992 and 1993 high school graduates)

Before Age/Funding Caps		*After Age/Funding Caps*	
13	of the grads were younger-arriving in Canada (initial placement was in junior high or elementary)	13/13	of the younger-arriving students required more than 3 years ESL support
31	of the grads were older-arriving in Canada (initial placement was in senior high)	15/31	of the older-arriving students were over age for high school attendance
		1/31	required more than 3 years ESL support
44	Total number of grads in '92, '93 (26% of total ESL population)	15	students would now be eligible to graduate (6% of total ESL population)

who were proficient at the advanced level in English on their arrival are among those who graduated in the class of 1992 or 1993. But these students would not be able to beat the newly proposed academic time clock in Alberta.

Interpreting the Findings

The overall dropout figures can be interpreted in a finer way. After investigating individual student progress course by course, we came to realize that there were three distinct categories of dropouts, which related in part to the quality of school experience. We hypothesized the following categories to describe what we observed as patterns among those who dropped out.

Dropout. After students had achieved successful progress for at least one year, they quit the educational system. This "active decision" could occur either after entering the mainstream, or during the supported integration stage. A number of reasons for dropping out were documented in the students' guidance files (those who participated in a withdrawal interview), but the most notable was the perception of impending failure and lack of interest and care on the part of their teachers.

Fall-out. These students are not "held in" by the educational system and they leave prior to any recordable progress. Generally, these students are at a beginner level and disappear from the school in the first or second semester. They tend to be students who have recently arrived and have limited L1 literacy and/or educational experience. Fall-out was not an active decision, as might be evidenced by participation in a withdrawal or exit interview with their guidance counsellor. Rather, it appeared to be the side effect of insufficient personal and educational support.

Push-out. Push-out refers to the removal of support services from students who still required them and/or were willing to continue with their successful progress, but for administrative reasons were either no longer identified or eligible. Once again, documentation in the students' guidance files supported categorizing students in this manner. Typically, push-out involves one of three factors: an unsuccessful or premature transition from ESL support to unsupported mainstream classes; a fixed length of educational support regardless of

individual need; or the high school attendance age cap, which has been established according to the educational norms of students who speak English as their first language.

It is push-out and fall-out that need to be most urgently addressed, and we return to this topic below.

From the size and scope of our study we feel confident for two reasons that these findings can be generalized to other groups of high school-aged ESL students across Alberta. First, our informal discussion with colleagues in ESL across Alberta tend to support the findings here. Our study is being duplicated in Calgary, in the same catchment area, in the Catholic high school. Initial results of the study lend firm support to our findings. Second, our comparative calculations of Alberta Education's 61% ESL dropout rate and the 74% rate that we report show the rates to be essentially similar. Their study included ESL students in junior high school, whose dropout rate is more comparable to that of advanced ESL high school students. Once this is factored into the sample, the hypothetical dropout for the remainder of the high school group in their study must be about 70%. This estimated rate for ESL students in 1986-1987 is similar to our overall rate and is even closer to our 1989 ESL dropout rate, suggesting that the dropout rate is actually worsening slightly. Given the size, scope, detail, and relation of findings between the two studies, we believe the findings reflect an accurate portrait of ESL dropout for the province.

Implications and Recommendations

This baseline study allows us to offer suggestions regarding tracking, policy, and administrative considerations.

A tracking study is only as good as the information on which it is based. Dirty data, and "noise in the system," lack of agreed-on definitions, and methodological and procedural considerations all add to the confusion. This is especially true for ESL students. Tracking will become particularly important in the future because in all probability the numbers of ESL students will increase in the system. Yet increasingly, they are liable to become lost in the system due to program cuts, the effect of push-out, lack of articulation between junior and senior high, as well as the philosophical move toward inclusiveness. In our view, *identification of ESL students should be based on need*, and once this is identified, a mechanism should be put in place to track these students throughout their schooling experiences. It would appear that these students remain forever at risk in the system, and thus tracking and monitoring their progress is crucial in addressing the overall dropout problem in the system.

From a policy and administrative perspective, it is push-out that needs to be most urgently addressed. We need to recognize that the concept of an age cap in high school discriminates against ESL students in a way that was not intended. It may have been designed to address perceived educational issues related to first language English students, but it adversely affects the progress of ESL students who are successfully progressing toward a high school diploma. Forcing these students out of high school produces an administrative barrier that will effectively end their educational participation. We need to permit ESL students who are successfully progressing with their education to continue in the same institution. We also need to recognize that language ability does not develop at the same rate for all students. Therefore, continued

funding and educational support should be available for those who do not achieve the language proficiency needed to be successful in the first three years of involvement in the school system.

Various kinds of support need to be made available: the one-size-fits-all approach to program design is not the answer. The demographics of the ESL population are shifting, and we must become more sensitive to the needs of the new arrivals. Significant numbers have limited first language literacy, and our tracking indicates that these students are not progressing beyond the beginner level. Basic literacy development, perhaps with the help of first language aides may be an alternative. For more advanced students, sheltered programs and adjunct courses, tutoring, and resource help may all offer the support they need to bridge more successfully into the mainstream.

Future Research

This study has three major limitations that require further research. One is its exclusively quantitative nature. It cannot supply information about how the ESL students perceived their educational experiences, nor can it inform us as to the educational expectations of ESL high school students. This more qualitative information is vital to completing our understanding of the statistical trends outlined in this study. The second limitation is inherent in the definition of the sample used for the study. An accurate picture of ESL student progress would require tracking the broader category of language minority students, including those who integrated early in their Canadian educational experience. Finally, future research must address the issue of the educational achievement for ESL students who have graduated from high school. An investigation of performance on provincial departmental examinations and on their progress in postsecondary institutions will complete the picture.

Conclusion

This study leads to the conclusion that access to instruction in English is the key to academic success for speakers of English as a second language. This access is a right, not a privilege. Students who have this right are defined in the Alberta School Act as: "those whose first language is not English and whose knowledge of English is insufficient to permit them to succeed in school and society" (Alberta Education, 1989).

Educational success is a combination of both progress and achievement, and the ESL dropout is a sign of how urgently we need to act if we are to promote the educational success of the present generation of ESL students in Alberta high schools.

References

Alberta Education. (1989). *Education programs, English as a second language*. Document Number: 02-03-03. Edmonton, AB: Author.

Alberta Education. (1992). *Achieving the vision report*. Edmonton, AB: Author.

Calgary Board of Education. (1991). *Retention study project*, Interim Report. Calgary, AB: Author.

Calgary Board of Education. (1993). *Student withdrawal report, 1991-1992 school year*. Calgary, AB: Author.

Collier, V. (1987). Age and rate of acquisition of second languages for academic purposes. *TESOL Quarterly 21*, 617-641.

Collier, V. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly 23*, 509-531.

Collier, V., & Thomas, W. (1989). How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Minority Students, 5*, 26-38.

Cummins, J. (1986). (Ed.). Language proficiency and academic achievement. In *Bilingualism in education, Aspects of theory, research and practice* (pp. 138-161). New York: Longman.

Early, M. (1993). Aspects of becoming an academically successful ESL Student. In B. Burnaby & A. Cumming (Eds.), *Socio-political aspects of ESL*. Toronto: OISE Press.

Employment and Immigration Canada. (1990). *Immigration and the population of Canada*. Ottawa: Author.

Grey, M. (1991). The context for marginal secondary ESL programs: Contributing factors and the need for further research. *Journal of Education Issues of Minority Students, 9*, 75-89.

Human Resources and Labor Canada. (1993). *Leaving school*. Ottawa: Author.

LeCompte, M. (1987). Can bad data produce good program planning? *Education and Urban Society, 193*, 250-268.

Minicucci, C. (1992). *Programs for secondary limited English proficient students: A California study*. Washington, DC: National Clearinghouse for Bilingual Education.

Olshtain, E. (1990). Factors predicting success in ESL among culturally different learners. *Language Learning, 40*(1), 23-42.

Radwanski, G. (1987). *Ontario study of the relevance of education, and the issue of dropouts*. Toronto: Ontario Ministry of Education.

Sinclair, R. (1992). Marginality, community, and the responsibility of educators for students who do not succeed in school. In Waxman, Herscholt et al. (Eds.) *Students at risk in at risk schools*. Corwin Press, CA.

Spener, D. (1988). Transitional bilingual education and the socialization of immigrants. *Harvard Educational Review, 58*, 133-153.

Statistics Canada. (1993, March 12). The daily. *1991 Census of Canada HIGHLIGHTS*. Ottawa: Author.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 297-318*

Ardra L. Cole
Ontario Institute for Studies in Education

Problems and Paradoxes in Beginning Teacher Support: Issues Concerning School Administrators

This article draws on information gathered in a qualitative research and development project focused on teacher induction and socialization in a school board in southwestern Ontario. I describe how principals, vice-principals, and department heads perceive and carry out their roles in relation to new teacher support and discuss several issues identified by them as problematic and paradoxical. In so doing I highlight some of the complexity associated with the administrator's role in school-based teacher development efforts and offer some suggestions for facilitating the resolution of some of the dilemmas they experience.

Cet article puise ses données à partir d'une recherche qualitative et d'un projet de développement qui visaient à étudier les processus d'induction et de socialisation des enseignant(e)s dans une commission scolaire située dans le coin sud-ouest de l'Ontario. Je décris de quelles façons les directeurs et les directrices d'écoles, les adjoints et les adjointes d'écoles, ainsi que les chefs de département perçoivent et jouent leur rôle respectif en ce qui concerne l'appui professionnel envers les nouveaux enseignants et les nouvelles enseignantes. Je discute également les aspects problématiques et paradoxales qu'ont identifié les administrateurs et les administratrices scolaires. En ce faisant, je souligne la complexité du rôle de ceux-ci et de celles-ci vis-à-vis les efforts de développement professionnel pour les débutants et les débutantes à partir de l'école-même et j'offre des suggestions pour faciliter la résolution de quelques dilemmes qu'ils/elles éprouvent.

> The nature of the person who goes into teaching is ... often that of someone who wants to help. And, as teachers and department heads, we have been helping students and colleagues. Then, we become a vice-principal or principal and that whole role changes. (Secondary School Principal)

Coming to terms with the dual role of supporter and evaluator of beginning teachers is one of the many challenges facing school administrators actively engaged in facilitating the development of beginning teachers. In this article I explore this and other such dilemmas school administrators struggle to resolve.

The data informing this writing are taken from information gathered in a qualitative research and development project focused on teacher induction and socialization in one school district in southwestern Ontario, Canada. This article reflects one project focus: the role of school administrators in the induction and support of beginning teachers. Other foci such as workplace relationships and school-based support programs are reported elsewhere (Cole, 1991a; 1991b; 1991c). In this article, I describe how principals, vice-principals, and department heads perceive and carry out their roles in relation to beginning

Ardra Cole is a professor in the Department of Applied Psychology. Her research and writing focus on teacher education and development with a particular emphasis on beginning teachers and qualitative approaches to educational research.

teacher support and discuss several issues identified by them as problematic and paradoxical. In so doing, I highlight the inherent complexity of the administrator's role in school-based teacher development efforts. Illustrations drawn from practices of four school administrators in different contexts provide some examples of how problems and paradoxes associated with new teacher support might be addressed. In conclusion, I offer some ideas regarding beginning teacher support and development based on perspectives and practical examples presented in the study and on what is known from relevant literature.

Background and Perspectives Informing the Work

As part of recent teacher education reform efforts, the attention of major interest groups in the education community—school districts, faculties and schools of education, teachers' federations, and government—has focused on the induction or settling in period for new teachers. The impetus for attention to new teachers has arisen partly out of recognition by those directly involved in teaching that teaching is becoming increasingly demanding and complex, and that new teachers, if they are to remain in the field, require specialized assistance and support in their early years of teaching. Calls for attention to new teachers have also come from others in the education community specifically interested in reforming teacher education. From this perspective, induction is seen as the *key to reform* with salutary effects on preservice and inservice teacher education and development (e.g., Fullan & Connelly, 1987). For a variety of reasons it is widely recognized that, perhaps now more than ever before, there is a need for change in the kind of entry experiences most new teachers have. Attempts to facilitate new teacher development through formalized support or induction programs have been widespread throughout North America.

Throughout almost every province in Canada, the setting for this article, the induction of new teachers is part of the reform agenda. In Ontario, for example, a 1991 survey of induction practices revealed that 89% of 109 school boards had some form of centrally sponsored orientation and/or induction program in place, and that 43% of those districts had system-sponsored induction programs combining a series of workshop activities with a formalized or informal school-based mentoring or support program (Cole & Watson, 1991; 1993). Increasingly recognition exists in practice and in the literature of the importance of school-based support for beginning teachers.[1] Much of the literature advocating school-based support efforts identifies, and in some cases defines, the roles of school administrators in new teacher induction and development (e.g., Anders, Centofante, & Orr, 1990; Burden, 1989; Cole, 1991a, 1991b; Cole & Watson, 1991; Crain & Young, 1990; DuFour & Eaker, 1981; Hetlinger, 1986; Hunt, 1968; Hunt & Associates, 1968; Leithwood, 1990, 1992). Also apparent in most of this work is the lack of formalized attention to those roles. The research on which this article is based was conducted in response to needs identified in the literature for a better understanding of the roles and responsibilities of school administrators and by those directly involved in beginning teacher support efforts.

Information Sources, Method, and Presentation

As indicated above, the information on which this writing is based represents one primary area of focus of a larger research and development project focused on new teacher development. For reasons cited above plus a dramatic increase

in hiring, teacher induction became a priority for the school board in which this project was situated. From 1989-1992 I worked closely with central office and school personnel in two main capacities. As a consultant I was involved in facilitating the design and development of a centrally sponsored district-wide assistance and support program for new teachers that had a strong school-based component. Emphasis was also placed on the role of school administrators in facilitating school-based support. As a researcher I followed the progress of program development (Cole, 1991a, 1991c) and explored integrally related issues such as beginning teachers' perceptions and experiences of their first year of teaching (Cole, 1991a, 1991b; Cole, Squire, & Cathers, in press; Cole & Knowles, 1992) and workplace relationships and teacher development (Cole, 1991b).

The program development aspect of the project was based on a model of participatory planning (Bach & Morley, 1987); the research component was broadly qualitative. Various methods of information gathering were used depending on the nature and purposes of the subproject inquiries. Contextual information was gathered in some cases through participation in project activities, in other cases through observations at project sites and through the use of the methods of *vignettes* and *prestructured cases* (Miles, 1990). Perceptions and experiences of various project participants (beginning teachers, experienced teachers, administrators, and members of the project planning team) were elicited through individual interviews and/or small-group discussions. All field observations were documented and all interviews and discussions were taperecorded and transcribed.

Information forming the basis for this article was gathered in two ways for two different but closely related purposes. Four principals volunteered their schools as pilot sites for the development of school-based induction programs. Over the course of two years, three principals (one secondary school and two elementary school) and one elementary school vice-principal focused attention on the provision of new teacher support in their schools with the overall intent of developing guidelines for other schools and school administrators in the district. Their participation involved, in part, a series of between six and eight individual in-depth interviews that focused broadly on their involvement in facilitating new teacher support and on issues, needs, and concerns arising from their efforts. Six secondary school department heads were also involved in group discussions about their roles.

As part of the district-wide focus on defining and supporting school administrators' involvement in new teacher induction, and in an effort to derive a broader-based understanding of how school administrators perceive their role in facilitating new teacher development, 23 elementary and secondary school principals and vice-principals volunteered to participate in one of four focus group discussions. In each two-hour discussion the participants considered the following topics in relation to new teacher support: their past and current involvement; perceptions of their roles; beliefs about and attitudes toward teacher development and beginning teacher support; and issues, needs, and concerns to be addressed over the short and long term. Taken together, the in-depth interviews and group discussions provide both close-up and wide-angle perspectives on the role of school administrators in facilitating new teacher development.

Using a method of analysis associated with a qualitative tradition as described, for example, by Eisner (1991) and Merriam (1988), themes and patterns were identified within each broad topic area and subsequently explored in the context of relevant literature on teacher induction, teacher education and development, and school improvement. An overarching theme emerging from all the information provided by school administrators was the sometimes paradoxical, sometimes problematic nature of their roles with respect to supporting new teachers.

Following an overview of a rationale for new teacher support, I present a brief summary of how the school administrators conceptualize their roles and describe their responsibilities associated with beginning teachers. I then turn to an exploration and discussion of the issues that represent to them problems and paradoxes. I follow with examples of how some school administrators have addressed some of these problems and paradoxes and conclude with recommendations for ways in which the full potential of teacher induction as a vehicle for teacher education reform might be realized.

Why New Teacher Support?

It has been said that teaching is the only profession in which beginners are catapulted into complex and demanding situations with no organized support system, yet are expected to perform with the fine-tuned expertise of their experienced colleagues. Beginning teachers are generally high achievers, academically well prepared, and keen to provide their students with stimulating and perfectly planned lessons and learning experiences. For the most part, though, their only practical preparation for such a complex and demanding task has been during the practicum component of the preservice program—a short-term, part-time assignment carried out with the assistance and close direction of an experienced teacher or teachers in charge of the classroom or classrooms where preservice teachers are placed. And, as any new teacher will admit, teaching in the practicum situation and teaching in one's own classroom are entirely different experiences.

Unfortunately, new teachers are often confronted with the least desirable teaching assignments their new school can offer and frequently are expected to assume teaching responsibilities for which they are not qualified (as in the case of a teacher prepared to teach at the intermediate level being assigned a position in the junior elementary grades). Ill-equipped classrooms in locations isolated from their colleagues, split grades, diverse teaching assignments that require multiple lesson preparation, and responsibility to teach many particularly challenging students are but a few of the realities with which new teachers are expected to cope (Cole & McNay, 1988; Huling-Austin, 1989, 1990; Odell, 1989; Varah, Theune, & Parker, 1986).

In addition to the primary task of taking direction of their new classroom, beginning teachers are expected, at the same time, to absorb the details of curriculum guides and school procedures, volunteer for extracurricular duties, and establish themselves in a school and community environment that is likely to be totally unfamiliar to them (Burke & Heideman, 1985). Accustomed to success and eager to prove their worth, new teachers struggle to manage all that they and others lay before them.

Although eager to push themselves, beginning teachers are often overwhelmed by pressures from all directions. They are dismayed by the number of students whose learning needs demand individual attention and frustrated by those whose behavior disrupts the classroom, throwing their carefully prepared lessons off course. Support from their colleagues may be haphazard or even nonexistent (e.g., Bullough, 1989; Bullough, Knowles, & Crow, 1991; Cole & Knowles, 1992).

As Cochrane-Smith and Lytle (1991) and Rosenholtz (1989) point out, it is likely that beginning teachers unfortunate enough to encounter a school with a sink-or-swim attitude toward newcomers will develop a fear of revealing what they perceive to be their inadequacies. Chances are they will start to isolate themselves even further from their peers, avoiding staffroom contacts and secretly dreading the official visit of the school administrator whose job it is to evaluate their performance (Zeichner & Gore, 1990). Little wonder, then, that so many talented individuals abandon the turbulent waters of the teaching profession in search of a calmer and more sheltered professional existence elsewhere. Research conducted in the United States shows that an estimated 30% of beginning teachers leave the profession during their first two years (Schlecty & Vance, 1983). In Canada a recent national survey indicates that at least 10% of beginning teachers find the job is not what they expected and leave within the first three years of qualifying (King & Peart, 1992).

The beginning experience is not always traumatic or short-lived, however. For some new teachers, the first year is a period of professional growth rather than a test of endurance (e.g., Cole, 1991b). The journey of beginning teachers embarking on their careers is nowadays often smoothed by induction programs: planned and formalized systems of assistance and support that enable novices not just to survive but to prosper during their first few years of teaching. And it is becoming increasingly evident that the kind of assistance and support most meaningful to new teachers, and of most benefit to other teaching professionals, takes place within the school setting where the wider professional development potential of induction programs can be more fully realized.

A school-based approach to new teacher support, however, places additional responsibilities and expectations on school administrators and staff. Principals, as designated school leaders, have not traditionally been involved in the induction process (Zeichner & Gore, 1990). And, as Leithwood (1990) claims, they may have unclear images of teacher development and how they might best facilitate it. The principals, vice-principals, and secondary school department heads involved in the research presented here recognized the importance of facilitating new teachers' induction to teaching and were interested in better understanding their role in such efforts. I turn now to a presentation and discussion of how this group conceptualized their roles and responsibilities and characterized their involvement in facilitating new teacher development.

The School Administrator and New Teacher Support

The Principal

> As a principal ... when you hire someone you have an interest in their development.... There is a responsibility there ... I am the educational leader of their

> school, so I have a responsibility for their development. (Secondary School Principal)

As designated school leader, the principal has a key role in initiating work toward the establishment and maintenance of a climate in which teacher support is valued as an integral part of teaching and professional development. Yet, for a variety of reasons, principals may or may not be directly involved in the provision of day-to-day support for beginning teachers. For example, many principals have responsibilities to the school board and community that regularly take them away from the school. They are not always readily available to provide "emergency care" or encouragement when it is most needed. Also, their evaluative role sometimes makes it difficult to establish a nurturing, collegial relationship with beginning teachers (a point to which I return below).

Generally, principals might be involved in the provision of new teacher support in the following ways:

- working to raise awareness among the experienced staff of the importance of new teacher support;
- initiating, facilitating, and orchestrating induction activities (e.g., organizing orientation activities, arranging classroom observations, introducing or matching beginning teachers with experienced resource or support teachers, arranging for release time for special induction activities);
- arranging, where possible, timetables and class assignments with consideration to new teachers;
- discussing with new teachers the direction of the school and their own philosophy of education and leadership;
- lending emotional support and encouragement to new teachers;
- providing information and advice; and
- offering overall cohesion and coherence to school-based support efforts.

The principals in this study attached a high level of importance to their facilitative role. They also described themselves as: liaison, friend, supporter, resource, mentor, guide, cheerleader, injector of humor, one who puts things into perspective and in balance, and one who adds a touch of realism to idealistic expectations and goals. The role of evaluator is a thorny issue for most principals because of the contradiction it presents when placed alongside some of their other roles. This is one of the paradoxical issues discussed below.

The Vice-Principal

> A vice-principal is there for support. When a new teacher comes [to you] and says, "Have you got a minute?" you get the coffee out, close the door, and cancel all your appointments because that is probably the most important time you'll spend in the day. (Secondary School Vice-Principal)

> [New teacher support] is the key to successful classroom teaching. It makes [new teachers] positive towards their job—if your job isn't pleasant, are you going to work at it very hard? ... If we want commitment from our staff we have to show them there's something worth committing to. [As an administrator] you're there to provide support so ... they want to come back, they want to give more to it. It's a small investment for a long-term pay-off. (Elementary School Vice-Principal)

Often it is the vice-principal who assumes primary responsibility for the provision of assistance and support to new teachers. Because for a variety of

reasons vice-principals usually have the most ongoing direct contact with new teachers, the roles of support lender and problem solver, formally or informally, have been added to their list of responsibilities. Vice-principals recognize the importance of this kind of responsibility and welcome opportunities to interact in myriad ways with beginning teachers. Designating teacher induction as a special responsibility not only reduces the chances of new teachers falling through the cracks but also gives an overall coherence and importance to teacher support and professional growth.

> You get to know [the new teachers]. You have lots of communication—even if it's spontaneous, social kinds of communication. [You] check with them often, even if you're just going down the hall. And while you're doing that, you're trying to analyze all the time what their needs are, what are their wishes (as far as involvement [goes]). [You] make yourself available to them on an informal basis so that when they have questions they're comfortable about that. You need to be approachable so that they don't feel they're imposing on you. (Elementary School Vice-Principal)

A number of vice-principals in the group discussions were relatively new to administration. Being new themselves and struggling to find their way in a new position and school, they empathized with beginning teachers. They recalled their own early classroom experiences and indicated an eagerness to provide whatever help they could to newcomers.

> I think we all remember our first year. I can't remember my second, third, or fourth years as well but ... I can still remember vividly some crucial days in my first year. That first year was really quite traumatic. (Elementary School Vice-Principal)

> I remember what it was like ... when I came in [to teaching] ... in times which were much gentler than they are today. I know the problems that I experienced, and I had very little fall-back support. Times are much more difficult [now] for new teachers and they need all the help and support we can give them. (Elementary School Vice-Principal)

> When I started [teaching] I quickly realized how isolating teaching can be.... You can be in your classroom, in the stockroom, [you can] go about your daily [practice] and not have much contact with others. It's really appreciated when a fellow staff member, vice-principal, or principal not only talks to you in the hall, but [now and again] pops into the classroom. I never did see it as a threat, I saw it as "Great! Here's someone to support me!" (Elementary School Vice-Principal)

The vice-principals repeatedly commented on the supportive nature of their role, how they regularly were called on to act as a buffer zone, and how often they tried to help beginning teachers put back into perspective issues and problems that had become distorted and overwhelming. In the kind of nurturing and support role that vice-principals saw themselves playing, the issue of evaluation also presented them with a dilemma. Like principals, they also have responsibilities related to teacher evaluation. And like principals, they experience a tension related to this role. Although this issue is discussed below, the following quote illustrates the point:

> When you're talking to [new teachers], you're still viewed as the vice-principal and [an] authority-type figure (although you don't maybe view yourself that

> way). You have to be sensitive to how they relate to you and be able to give them the feeling that they can come to you for help. [You don't want them to take what you say] as the definitive law or rule [of] the school. (Elementary School Vice-Principal)

The Department Head

In the secondary school, it is the department head who perhaps assumes the most critical role in terms of the actual delivery of, and direct involvement in, induction activities. This is due to the structure of the secondary school, the department head's responsibilities to facilitate professional growth, and because part of the department head's timetable is set aside for facilitative or coordinating responsibilities. A secondary school principal describes the role of the department head in this way:

> The role of the department head is to make certain that the prescribed curriculum is understood by the teacher, is delivered by the teacher on schedule, and that the new teacher's work is coordinated with other people in the department.... So, although the department head's main responsibility would be to curriculum, I would expect [it] to go beyond this to [include supporting] the new teacher induction team [at the school level]. (Secondary School Principal)

Like many of the vice-principals, the department heads referred to their own early experiences to explain their interest in and commitment to new teacher support. For example, one explained her position as follows:

> This is the first time I've had a new teacher in my department. Remembering my history [with other] boards, that I didn't get the support that I would like to have had ... I felt particularly keen on making myself available to [the new teacher].... I said to myself, "I don't want [what happened to me] to happen to her." ... It was really important that she know that I am available and I'm there to support her and help her whenever I can. (Secondary School Department Head)

Department heads in one school identified the following ways in which they were involved in new teacher support:

- creating a climate of support within the department, and encouraging attitudes and practices of helping and sharing;
- being available by telephone for after-hours support and for consultation over the summer before classes start;
- orienting new teachers to the school and department layout, policy, procedures, routines, and the like;
- offering daily, immediate support (often of a technical, detailed, or subject-specific nature);
- coordinating inter- or intradepartment work groups;
- providing coverage to allow teachers to visit other classrooms;
- coordinating schedules to allow time to meet regularly with new teachers;
- requesting that new teachers' classes be located close to those of helpful experienced colleagues; and
- making appropriate class assignments and, where possible, allocating a reduced workload.

Problems and Paradoxes

Despite an obvious interest in and commitment to new teacher support, school administrators have a number of concerns relating to teacher induction. Issues

that repeatedly emerged in our conversations reflect certain realities associated with new teacher support and appear to be ongoing sources of tension and concern for school administrators. The issues are paradoxical in nature in that they represent conflicts or contradictions. School administrators experience conflict associated with their role in new teacher support and with their relationships with new teachers as they struggle with the dual role of evaluator and support, strive to find an appropriate balance between facilitation and direction, and as they try to be respectful of teachers' work style preferences in light of their own. They struggle to provide facilitative work contexts for new teachers while maintaining their commitment to other more experienced teachers in the school. They are often frustrated by contradictions between new teachers' needs and capabilities and the needs and structures of the school and school system. And they are faced with the dilemma of being a primary support provider while receiving little or no assistance or support in that role. In this section, I present and discuss each of these problems and paradoxes in the context of relevant literature. Following this I give examples of how some administrators have addressed them.

Conflicts Associated with School Administrators' Roles and Relationships with New Teachers

Assistance versus Assessment

The comment made by the secondary school principal at the beginning of this article reflects one of the most challenging dilemmas facing school administrators: coming to terms with the dual role of supporter and evaluator of beginning teachers. Some administrators find it difficult to accept the reality and implications of being perceived as the supervisor or evaluator when they prefer to be seen in a helping role. One principal illustrates:

> The assistance I thought I was providing the new teachers sometimes was not being picked up as that. They saw me as the person who was going to observe them to determine whether they will be given a permanent contract even though I thought I was working with them on a supportive level. (Secondary School Principal)

Related to this is the challenge of trying to be a helper and evaluator simultaneously. A vice-principal explains:

> We have to see our new teachers [for evaluation purposes] twice before November 30th, and that's a really threatening situation for them.... It's really important that we don't come in just as administrators.... [Although] we're evaluating, we [also] are there for their growth. [It's important that they understand that] as teachers starting in, their first year will be full of opportunities for growth. (Elementary School Vice-Principal)

An elementary school principal sums up the reality of the situation in a single statement: "Once you are on the other side of the desk there is no denying it, you *are* the boss."

The assistance-assessment dilemma articulated by these administrators is well recognized in the literature. Induction programs based on an assistance model acknowledge the complexity of teaching and the individuality of the teacher. In these programs professional development is viewed as an ongoing process facilitated through self-assessment and reflective practice and with

attention to personal and professional support. Evaluation, other than self-assessment and constructive or formative feedback, is not part of this kind of induction process. (New teacher support in this school district and in most Canadian schools is based on an assistance model.)

Teachers who know they are being evaluated or who see their helper also as an evaluator are less likely to take risks, ask questions, or seek help. As Scriven (1988) maintains:

> Without this separation [of assistance and assessment], it is unreasonable to expect teachers to go to formative advisors about their weaknesses. One might as well expect clients to seek advice from attorneys who are doubling as judges on the same case. (pp. 114-115)

Responsibilities associated with certification or contractual matters are typically assumed by those not directly involved in the provision of induction support and assistance, either mentor or support teachers (Neal, 1992; Odell, 1990) or in some cases university teacher educators (McEvoy & Morehead, 1987). Neal (1992) points out that the term *evaluator* or adjectives associated with that role are not typically used to characterize the functions of a mentor or support person. Terms such as *supporter, confidant, guide, advisor,* and *encourager,* which all assume the existence of a level of trust and safety, more often characterize the relationship between a novice and experienced teacher in a supportive role.

The problem for the principals and vice-principals in this study is that the evaluative role they must perform as part of their administrative responsibilities is at odds with the support role they wish to assume. They stand, according to how Neal (1992) distinguishes these roles, with one foot at either end of a continuum of teacher development activities. Goodlad (1984) argues that as much as they would like it to be otherwise, principals cannot be both an instructional leader and an evaluator. This is precisely the dilemma faced by this group of principals and vice-principals. And for them, and others like them, it is a dilemma for which there is no easy resolution.

Treading the Fine Line between Development and Intervention

The problem of judging when to intervene for the benefit of the new teacher, the students, and the school presents another challenge.

> I'm always in a quandary about how to give the [new teachers] enough room to establish themselves and do the things they want to do, and not give them the feeling that I'm standing over top of them. I want to be able to guide them but I don't want it to be a "sink or swim" situation. It's difficult to always keep an eye on things and know when it is time to give a little assistance or advice and when to let [something] go a bit longer to see what happens. (Elementary School Vice-Principal)

This statement could be made by anyone in a facilitative role. It well captures the ongoing dilemma inherent in any experiential learning situation. In the beginning teacher development literature, the pressure-support tension is often articulated in relation to mentoring roles (Head, Reiman, & Thies-Sprinthall, 1992; Shulman & Colbert, 1987; Wildman, Magliaro, Niles, & Niles, 1992). This tension is exacerbated, however, when the person assuming a facilitative function also has wider accountability concerns. As this principal indicates,

there can be an urgency to a situation where immediate and direct intervention is required:

> In a situation where the [beginning teacher] is really struggling, you have to get a plan of action in place fairly quickly because if something is not done then the students are suffering. How much they are suffering will determine how quickly you have to move. So, as a result of that type of role ... certainly the principal is not going to be seen as the friendly person who has come in to help out. (Secondary School Principal)

An elementary school principal further illustrates how difficult it is to walk the line between facilitating development through experience and using direct intervention:

> [One beginning teacher] didn't want to take my advice. I would give her suggestions of pitfalls that were lying out there but she continued to follow on down the road and, as a result, after about two months in the job she was in a lot of trouble.... I found it extremely challenging and a major dilemma. (Elementary School Principal)

One elementary school vice-principal identified a role for the beginning teacher in addressing the issue. She reflects on the importance of communication between administrators and beginning teachers and on the role of beginning teachers in determining the nature and extent of the help and support they need:

> The first year is a delicate year. You don't want to treat [the beginning teachers] like babies; they are professionals. Some of them want to learn from experience and not always have someone there to grab them before they fall flat on their face.... Sometimes it's hard to know when to let them alone and when to help. It's critical to set up an honest communication system so that they can help you [gauge that]. (Elementary School Vice-Principal)

According to this administrator, it is important to develop a relationship with beginning teachers based on mutual trust and respect. Given the hierarchical nature of the relationship between administrators and beginning teachers, and the especially vulnerable position in which beginning teachers are placed because of their probationary status, the development of such a relationship is another challenging and complex process.

The development of collegial relationships, both among teachers and between teachers and administrators, is as much of an organizational issue as an interpersonal one as the literature on school culture indicates (e.g., Heckman, 1987; Johnson, 1990; Little, 1990; Rosenholtz, 1989). The creation of a work context conducive to help seeking and risk taking is required so that beginning teachers feel sufficiently secure to be open and honest with their superiors. Principals (and other administrators) have an important role in the creation of such a context (Barth, 1990; Cole, 1991a, 1991b; Fullan, 1988; Fullan & Stiegelbauer, 1991). This vice-principal describes the challenge and goal of mutual acceptance and support:

> We need to accept the fact that a new teacher, any teacher, is always learning. We need to both support one another and accept support. Once we can accept support we're more apt to give support. (Elementary School Vice-Principal)

Encouraging Help Seeking and Collaboration While Respecting Individualism

> There are different types of new teachers. Some are very quiet and feel that by coming to you they are stating that they are having difficulty when they're not. You have to approach them and ask about specific things otherwise they won't say anything to you. Others are more vocal and will come right out and say that they are having problems. You have to know the personality types of the teachers and work with the different types [accordingly]. (Elementary School Vice-Principal)

> Teachers come into the profession with different sensitivity levels. We really need to practice with our staff members what we ask them to do with our kids: lots of praise, positive reinforcement, encouragement, being there to listen, being honest. (Elementary School Vice-Principal)

Administrators recognize the need to be respectful of sensitivities and responsive to teachers' individual differences; to different personalities and working styles; to the teacher who prefers to work alone in the classroom; to the teacher who appears not to heed advice; and to the quiet or shy teacher. At the same time, they are mindful that beginning teachers may hesitate about asking for help or indicating that they need it, or that they may cover up problems or concerns feeling that to admit them would be to reveal inadequacies (Cole & Knowles, 1992; Neal, 1992). New teachers are even more likely to insulate themselves from their superiors (Zeichner & Tabachnick, 1985).

It is often the case as well that the most pressing problems are those for which beginning teachers are most hesitant to seek advice. A vice-principal comments on this issue with respect to classroom discipline, an area identified as a significant concern for beginning teachers (Cole et al., in press; Odell, 1989; Veenman, 1984):

> Some new teachers ... may need more support than others but for all of them there's a point in the year when [discipline] is a problem. This is a stressful time [for new teachers].... But the big problem is that they feel they're being "marked down" if they send kids to the office. They feel that's in some way a failure on their part. [I] take a lot of time to say, "Look, we've all been through the first year. It's fine. That's what I'm here for. We'll help you in any way we can. (Secondary School Vice-Principal)

So school administrators are faced with the problem of how to communicate a climate of support and encourage beginning teachers to ask questions and seek help without appearing to criticize or interfere with certain preferred work styles and behavior. Similarly, they need to be careful not to discourage those who are more openly inquisitive and interested in engaging in collegial relationships.

Conflicts Associated with Responding to the Professional Development Needs of All Teaching Staff

Principals are mindful of their role as staff developers. Also aware of the special needs of new teachers, they struggle to attend appropriately to new teachers without jeopardizing the development and professional well-being of others in the school. They face a number of problems related to this dilemma, some structural and some attitudinal. For example, they are aware of the need to find ways to avoid overburdening experienced teachers with mentoring

responsibilities or with tasks that have not been assigned to new teachers in an attempt to lighten their workload. Even though they are aware that it is both within their authority and of benefit to beginning teachers to make structural modifications in class assignments, workload, and extracurricular responsibilities, they are also aware that such modifications may present problems for others on staff. Two principals express their predicament:

> At whose expense can the new teacher be given a particular timetable [for example]? ... There are other teachers who need to have good experiences in the classroom as well.... Everyone needs support from first year to 97th. (Secondary School Principal)

> Sometimes [by asking experienced teachers to help out] you feel you're putting more responsibility on those ... who already have a fair bit on their plate [but] you know that they have dealt with certain students and have a good handle on how best to work with them. (Elementary School Principal)

What often happens, the principals suggest, is that they end up assuming many of the added responsibilities in order to maintain a level of fairness. As one elementary principal remarked, "You can't force teachers to do anything, you can only set the tone and an example. You have to be careful not to wear out the older teachers and not to seem to show favoritism."

The principals also have to deal with attitudinal barriers to new teacher support put up by some teachers in their schools. Without appearing to show favoritism, they work hard at encouraging those resistant to helping and at instilling a spirit of collegiality. It is also difficult to work with those exhibiting the "I survived it" syndrome, those who see the first year of teaching as a rite of passage. This principal describes the sentiments of some experienced teachers toward induction activities:

> [Now that] we have more people involved [in new teacher support] and more opportunities in the school year to talk about induction activities ... there is a degree of envy out there among some of the teachers who have more than eight years experience, who do not have positions of responsibility, and who do not know the facilities and resources of this system as well as do the new teachers. (Secondary School Principal)

Again, the administrators are faced with a dilemma embedded in issues of school culture. Although they can make the kind of structural provisions for new teachers suggested in the literature (Cole, 1990a; Cole et al., in press; Huling-Austin, 1989), unless they are able to overcome attitudinal barriers and are successful in fostering norms of collegiality, such well-intended actions may be misconstrued or even backfire. As Johnson (1990) observes, "Even when the schedule is right and time is sufficient, distrust, disrespect, and dissension can undermine collaboration" (p. 178). In contrast is the kind of work context described by Gherke (1987) where "a group of teachers, administrators, and support staff members in a school who have been helped themselves to such a degree and in such ways that they will see helping as an inherent part of their roles with newcomers and oldtimers" (p. 110). In this latter kind of environment, fairness and competition for favor and resources are non-issues.

Contradictions Between New Teachers' Needs and Capabilities and the Needs and Structures of the School and School System

Assisting and Supporting New Teachers Within the Bureaucratic Structures of School Systems

Related to the problems identified above of overcoming structural and attitudinal barriers within schools are similar problems reflective of the broader bureaucratic structure of school systems. The administrators in the study identified some major obstacles to providing more satisfying induction experiences for new teachers that they believe to be beyond their control and influence.

In order to provide opportunities for new teachers to become oriented to the school and classroom and to adequately prepare for their first assignment, and in order for class assignments and schedules to be made with new teachers in mind, hiring should take place sufficiently early. More often than not, vacancies and new appointments are made just prior to or even after the beginning of the school year. This presents obvious problems for both beginning teachers and school administrators because neither has any time to prepare. A secondary school principal expresses his frustration:

> Somehow we have to find a way to provide these [new] teachers with a lightened workload so that things are more manageable for them.... [One new teacher at my school] inherited a horrendous timetable. By the time everything had trickled down and somebody finally said, "I have a position here," [that] position had a very difficult timetable.... If we could identify our new people in time so that we could provide them with a reasonable timetable, I think that would be beneficial for them. (Secondary School Principal)

Policies and practices of hiring new teachers mid-year and on temporary, limited contracts present administrators with additional problems. Also, they often are forced to assume a more directive and reactive, rather than supportive and facilitative, role with these beginning teachers.

Adequacy of financial resources to support induction activities is another issue. Recent budget cutbacks as a result of a recent economic recession have had an impact on options for new teacher support activities. With little money available for additional resources or release time to allow beginning teachers to engage in induction activities such as observing in other teachers' classrooms, administrators either have to assume the responsibility themselves or not provide such professional development opportunities. These are what Fullan and Stiegelbauer (1991) call classical organizational dilemmas associated with the middle management position of principals who are always trying to best serve both the teachers in their school and the senior administrators in central office.

New Teachers' Preparedness to Teach and the Needs of the School

Finally, new teachers' level of preparedness for the realities of the classroom is a concern that continues to plague administrators and new teachers alike. Administrators question both the initial selection practices by which prospective teachers are admitted to preservice programs, and the focus and content of the preparation programs themselves:

> With a greater emphasis on academic preparation for colleges of education, we're missing the candidates who would be able to come into the classroom as

> more rounded teachers. The [new teachers] have lots of energy and strong academic backgrounds but have a hard time interacting with the students and relating to those who experience academic difficulties. I would like to see the new teachers with more of a sense that they are working with young people than [that] they're teaching a subject. My concern is that so much of their preservice training is academic and so little of it is based on relationships. (Secondary School Principal)

These are not new concerns. They have been voiced repeatedly in teacher education reform literature across the country and beyond (e.g., Bowman, 1991; Fullan & Connelly, 1987; New Brunswick Commission on Excellence in Education, 1992; Tuinman & Brayne, 1988) and are currently at the heart of teacher education reform efforts in Ontario and elsewhere. Nevertheless, because reform is slow and incredibly complex, it will take time for efforts to be realized at the level of the school. In the meantime, school administrators continue to hire and work with teachers who need more guidance than they feel should be required.

It is also clear from their comments that these administrators recognize that the kind of change required must reflect the shared interest and responsibility of both faculties of education and schools:

> We need closer contact with our faculties of education to make sure that people coming out have a more realistic picture of what teaching is all about. [Teaching] has changed in the last few years since we started. We need that kind of contact so that we're setting up [new teachers] for success, not failure. (Elementary School Vice-Principal)

Helping While Needing Help

It should come as no surprise that school administrators also feel the need for guidance and support as they assume and carry out responsibilities related to teacher development and school improvement. The following are some specific needs expressed by the school administrators in this study:

- receiving feedback from new teachers about the support offered them;
- having a clear idea of the school district's position and mandate regarding support for beginning teachers;
- participating in inservice professional development activities in specific areas (e.g., documenting teacher performance and assisting with professional growth plans); and
- having opportunities to meet with other administrators to discuss issues related to teacher induction.

These concerns are echoed and discussed in literature on the principalship. For example, Blumberg and Greenfield (1980), Fullan and Stiegelbauer (1991), and Sarason (1971) comment on the loneliness of the principal's role. Fullan and Stiegelbauer (1991) also point out that usually principals receive little assistance from central administration with respect to the implementation of new programs.

Often school administrators are victims of the kind of work conditions that they are striving to change for the teachers in their schools. And, like new teachers, they cover up or hesitate to share their fears and concerns, and have little chance but to sink or swim:

> When we were new [vice-principals] it was not much different than it is for new teachers. Remember when we all got together that first time? "How's everything?" "Fine. No problems." And then, on the way home in the car, you thought, "Oh, if they only knew." (Elementary School Vice-Principal)

Toward Resolution

The paradoxical nature of the issues and concerns identified by the school administrators in this study reflect certain realities associated with new teacher support and development. An acknowledgment and discussion of the problems and concerns they identify serve to highlight the complex nature of teacher education and development and the need to find productive ways of engaging all partners in the process of teacher education reform. And although there are no easy or fast solutions to these persistent problems, focused attention on facilitating new teacher support in their respective contexts afforded four school administrators opportunities to better understand and deal with some of the complexities associated with facilitating beginning teacher development in the workplace. The four administrators involved in the pilot projects for the development of school-based support programs provide important insights into some of the issues.

Learning from Four School Administrators

The problems and paradoxes discussed above were issues for all four administrators; however, the extent of the concerns varied across contexts and individuals, and mostly lessened over time. School culture and leadership style were critical factors contributing to and/or helping to resolve most of the dilemmas, particularly those associated with evaluation, intervention, fairness, and sensitivity to individual differences. For these four administrators, facilitating new teacher support was (and is) integrally connected to the broader issue of school-wide professional development. Providing beginning teachers with a facilitative induction to teaching in all cases meant significant changes in the schools. In some cases the changes involved reallocation of space, time, and resources; in others changes represented challenges to traditional isolationist norms and patterns of professional interaction. This latter kind of change seems fundamental to the realization of the broad potential of teacher induction and to diminishing the problematic and paradoxical nature of school administrators' roles.

Three of the administrators explicitly addressed the tension created by the dual role of helper and evaluator. Two focused attention on reconceptualizing evaluation practices and their role in the process by placing emphasis on ongoing teacher-directed development and by increasing their amount of *informal* time spent in classrooms. Both administrators described the importance of their presence in the classroom being seen as *natural*. One indicated her belief that given time, opportunity, and appropriate emotional support, [all] teachers would feel freer to identify and attend to their own professional development needs. The third administrator, for whom the evaluator/support tension provided difficulties, chose to accept the evaluator role and shift primary responsibility for various kinds of ongoing support to others in the school "much better placed to assume a supportive role and respond to [beginning teachers'] needs."

Involving others in support roles often means a change in professional interaction patterns and norms. All four administrators stressed the importance of modeling and encouraging collegiality and working together. For some, this meant challenging well-established norms of individualism, a task defined by one as "like pulling teeth." Building a culture of community in which people naturally work together, support one another, and share ideas and responsibilities is a major facilitator of beginning teacher development with significant benefits for all involved (Cole, 1991b). The four administrators involved in developing school-based support programs spent considerable time creating formal and informal opportunities for teachers (new and experienced) to work together.

For example, in one school classrooms were clustered according to grade level or subject area to afford better opportunities for teachers to work together and help one another. Also, opportunities for teachers to informally exchange ideas were provided at regularly arranged lunchtime and after-school conversation sessions. Holding staff meetings in different classrooms on a rotating basis and devoting time at those meetings to sharing information and ideas from any professional development activity or event in which a teacher might be involved are other ways administrators have tried to encourage more professional interaction.

Another principal whose efforts to "lessen isolationism" met with considerable initial resistance created "nonnegotiable" opportunities for teachers to observe in other classrooms within the school. Even though this meant that she had to cover the visiting teacher's class, she believes that teachers need to experience the benefits of working with others and that it is her responsibility as principal to create opportunities for those benefits to be realized. Over time this principal came to define her role in relation to new teacher support as encouraging (new and experienced) teachers to work together and support one another rather than assuming full and direct responsibility for the provision of new teacher assistance and support herself (as she initially did).

One administrator gave an example of how efforts to encourage joint work among teachers unexpectedly paid off. At the request of some experienced teachers, a series of professional development activities initially arranged for beginning teachers in the school was expanded to include all staff interested in attending.

> [Over the course of the sessions] one of our most experienced grade 1 teachers latched on to one of our beginning grade 1 teachers. When it came time [later on] to set out their goals for professional opportunities for growth [a district-sponsored individualized professional development plan], she chose the beginning teacher as her mentor! It was a shock to me although I didn't let on. [When we talked about it], she said, "I just wanted to take advantage of all these new ideas fresh out of teachers' college and the enthusiasm this person has to share." (Elementary School Principal)

Those administrators working in schools not used to high levels of professional collegiality and collaborative work acknowledge that change in interaction patterns is slow to develop but worth continual encouragement. They see an emphasis on new teacher support as one important vehicle for such change. As one principal remarked, "There is greater awareness among other staff

people that we do have a greater responsibility to help new teachers get established." This greater awareness and acceptance combined with a gradual increase in the number of new teachers who themselves have had supportive induction experiences contribute over the long term to a reorientation of professional workplace norms.

When helping, sharing, supporting, and working together become the accepted way of doing things, that is, the natural way for people to relate to and work with one another, many of the dilemmas identified by the school administrators in this study become non-issues.

Moving Ahead

I conclude with a summary of the dilemmas presented in this article and some suggestions that, if given further consideration, might move us closer toward resolution.

Conflicts associated with school administrators' roles in new teacher support and their relationships with new teachers. Traditionally, teaching has been characterized by norms of isolation and evaluation. For a variety of reasons, many new (and not so new) teachers are reluctant to ask questions or seek advice about teaching; however, school administrators are increasingly encouraging more collegial and collaborative practices. Teachers are unlikely to sincerely and willingly respond to such encouragement unless:

- appropriate attention is paid to prevailing leadership norms and practices and established patterns of professional interaction both in schools and in the profession at large;
- efforts are made to rethink evaluation policies and practices at both school and district levels so that they are more holistic and growth-oriented and involve both teachers and administrators in a process of shared responsibility and decision making; and
- a focus is placed on building work and learning environments that encourage rather than dissuade risk taking.

Conflicts associated with responding to new teachers' needs while maintaining commitment to all staff. As indicated above, many of the staff-related dilemmas associated with new teacher support are naturally resolved when workplace norms are reoriented toward helping, sharing, supporting, and working together. School administrators and teachers need to focus on:

- creating a context in which teacher support and development are a natural part of school practice so that new teacher support does not appear as an aberration and is not seen as creating additional responsibilities for teachers and administrators.

Contradictions associated with new teachers' needs and capabilities and the needs and structures of the school and school system. The current bifurcated system of teacher education is based on a limited view of teacher education and development that lacks conceptual and practical continuity from preservice to inservice teacher education and development. It is set up to expect an unreasonably high level of sustained performance from even experienced teachers, and new teachers introduced to the system have the same high expectations thrust upon them. If dilemmas associated with administrators' perceptions of new teachers' lack of preparedness to teach are to be resolved, concerted effort is required by schools and faculties of education to:

- develop shared understandings and goals in relation to preservice and inservice teacher education; and to
- clarify the roles that each institution might appropriately assume at various points along a career-long continuum of teacher education and development.

The administrative and bureaucratic structures of school systems often militate against innovation or change. Such is the case with new teacher support and development. To overcome obstacles to teacher development created by the bureaucratic structure of school systems requires:

- a strengthening of relationships between schools and school board central offices;
- a more holistic vision of the roles of various institutions in ongoing professional development;
- the development of a set of mutual goals reflecting a commitment to teacher development and school improvement; and
- greater responsibility for decision making to be placed with teachers and school people.

The dilemma of being a primary support provider while receiving little or no assistance and support in that role. Attention to beginning teachers has been called the "key to reform" (e.g., Fullan & Connelly, 1987; Fullan & Stiegelbauer, 1991). There is little doubt that providing facilitative experiences for beginning teachers is a good thing. The administrators in this study, like many others, are committed to the idea of beginning teacher support and are actively involved in facilitating various induction opportunities. Yet many are frustrated by the dilemmas they face and the dim prospect of their resolution.

As school districts continue to move toward site-based management and as increased emphasis is placed on school-university partnerships, more and more responsibility for initial and ongoing teacher education is falling on the shoulders of school administrators. For the most part, administrators are expected to assume that responsibility with little or no guidance and support, and with minimal opportunity to acquire knowledge of the rationale for or long-term implications of such school-based initiatives. Because substantive change takes time and commitment, and commitment itself requires time to develop, school administrators as overburdened middle managers are able to do little more than try to implement policies and practices and struggle to resolve associated dilemmas.

Complex issues such as those identified here need to be addressed if substantive change is indeed to take place; however, without appropriate levels of support and involvement to allow such consideration, reform efforts such as those focusing on beginning teachers are likely to remain "just some more good ideas."

Note

1. A plethora of literature supports the notion of school-based induction support. More specifically, rationale for this kind of support can be found in four areas of educational literature: work dealing generally with the principles and practices of teacher induction (e.g., Andrews, 1986, 1987; Brooks, 1987; Cole, 1990b; Cole & McNay, 1988; Griffin & Millies, 1987; Huling-Austin, 1990; Huling-Austin, Odell, Ishler, Kay, & Edelfelt, 1989); literature linking teacher development and school improvement (e.g., Fullan, 1985, 1991; Fullan & Stiegelbauer, 1991; Lieberman, 1986; Thiessen & Kilcher, 1991; Wideen & Andrews, 1987); work focused on

school context and culture (e.g., Cochrane-Smith & Lytle, 1991; Fullan, 1990; Fullan & Hargreaves, 1991; Heckman, 1987; Little, 1987; Rosenholtz, 1987, 1989; Sarason, 1982); and research on mentoring (e.g., Bey & Holmes, 1990, 1992; Kilcher, 1989; Little, 1990; Little & Nelson, 1990; Zimpher & Rieger, 1988).

References

Anders, L.W., Centofante, E.F., & Orr, J.T. (1990). New teachers: Easing the transition. *NASSP Bulletin, 70*(1), 51-52.

Andrews, I.H. (1986, April). *An investigation of the academic paradigms underlying induction programmes in five countries.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Andrews, I.H. (1987). Induction programs: Staff development opportunities for beginning and experienced teachers. In M.F. Wideen & I. Andrews (Eds.), *Staff development for school improvement* (pp. 142-153). PA: Falmer Press.

Bach M., & Morley, D. (1987). Participatory planning: An alternative conceptual framework and methodology. In M. Bach & D. Morley (Eds.), *A social learning approach to participatory planning.* Toronto, ON: Faculty of Environmental Studies, York University.

Barth, R. (1990). *Improving schools from within.* San Francisco, CA: Jossey-Bass.

Bey, T.M., & Holmes, C.T. (Eds.). (1990). *Mentoring: Developing successful new teachers.* Reston, VA: Association of Teacher Educators.

Bey, T.M., & Holmes, C.T. (Eds.). (1992). *Mentoring: Contemporary principles and issues.* Reston, VA: Association of Teacher Educators.

Blumberg, A., & Greenfield, W. (1980). *The effective principal.* Boston, MA: Allyn and Bacon.

Bowman, J. (1991). *Report to the College of Teachers on teacher education in British Columbia.* Vancouver, BC: BC College of Teachers.

Brooks, D.M. (Ed.). (1987). *Teacher induction: A new beginning.* Reston, VA: Association of Teacher Educators.

Bullough, R.V., Jr. (1989). *First year teacher: A case study.* New York: Teachers College Press.

Bullough, R.V., Jr., Knowles, J.G., & Crow, N.A. (1991). *Emerging as a teacher.* London, UK: Routledge.

Burden, P.R. (Ed.). (1989). *Principals as staff developers, Journal of Staff Development* (Theme issue), *10*(1).

Burke, P., & Heideman, R. (Eds.). (1985). *Career-long teacher education.* Springfield, IL: Thomas.

Cochrane-Smith, M., & Lytle, S.L. (1991, April). *Telling a different story: Teacher research as research on teaching.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Cole, A.L. (1990a). Personal theories of teaching: Development in the formative years. *Alberta Journal of Educational Research, 36,* 203-222.

Cole, A.L. (1990b). Helping teachers become real: Opportunities in teacher induction. *Journal of Staff Development, 11*(4), 6-10.

Cole, A.L. (1991a). *Four schools, four ways of doing things: A participatory approach to the development of school-based induction programs.* Paper presented at the Annual Conference of the Canadian Society for the Study of Education, Kingston, ON.

Cole, A.L. (1991b). Relationships in the workplace: Doing what comes naturally? *Teaching and Teacher Education, 7*(5/6), 415-426.

Cole, A.L. (1991c). *Teacher induction: Opportunities for career-long and school-wide professional development.* Toronto, ON: OISE Press.

Cole, A.L., & Knowles, J.G. (1992). *Beginning teachers talk: Who listens? Who learns?* Paper presented at the Annual Conference of the Canadian Society for the Study of Education, Charlottetown.

Cole, A.L., & McNay, M. (1988). Induction programs in Ontario schools: Issues and possibilities. *Education Canada, 28*(4), 4-11, 44-45.

Cole, A.L., Squire, F.A., & Cathers, P. (in press). *Supporting beginning teachers: A handbook for school administrators.* Unpublished report, Ontario Institute for Studies in Education

Cole, A., & Watson, N. (1991). *Support for beginning teachers: Ontario perspectives.* Report to the Teacher Education Council, ON.

Cole, A.L., & Watson, N. (1993). Beginning teacher development in Ontario: The ebb and flow of policy and practice. *Journal of Education Policy, 8*(3), 241-255.

Crain, C.L., & Young, T.A. (1990). *Tips for principals.* Reston, VA: National Association of Secondary School Principals.

DuFour, R., & Eaker, R. (1981, September). The principal as leader: Two major responsibilities. *NASSP Bulletin, 71*(500), 80-89.

Eisner, E.W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Macmillan.

Fullan, M.G. (1985). Change process and strategies at the local level. *Elementary School Journal, 84*, 391-420.

Fullan, M.G. (1988). *What's worth fighting for in the principalship? Strategies for taking charge in the elementary school principalship*. Toronto, ON: Ontario Public School Teachers' Federation.

Fullan, M.G. (1990). Staff development, innovation, and institutional development. In B. Joyce (Ed.), *Changing school culture through staff development* (pp. 3-25). Alexandria, VA: Association for Supervision and Curriculum Development.

Fullan, M.G. (1991). *Productive educational change*. Lewes, UK: Falmer Press.

Fullan, M.G., & Connelly, F.M. (1987). *Teacher education in Ontario: Current practice and options for the future*. Report to the Teacher Education Review Steering Committee, Toronto, ON.

Fullan, M.G., & Hargreaves, A. (1991). *What's worth fighting for?: Working together for your school*. Toronto, ON: Ontario Public School Teachers' Federation.

Fullan, M.G., & Stiegelbauer, S. (1991). *The new meaning of educational change*. Toronto, ON & New York: OISE Press & Teachers College Press.

Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.

Gherke, N.J. (1987). On helping the beginning teacher. In G.A. Griffin & S. Millies (Eds.), *The first year of teaching: Background papers and a proposal* (pp. 105-114). Chicago: University of Illinois-Chicago.

Griffin, G.A., & Millies, S. (Eds.). (1987). *The first year of teaching: Background papers and a proposal*. Chicago: University of Illinois-Chicago.

Head, F.A., Reiman, A.J., & Thies-Sprinthall, L. (1992). The reality of mentoring: Complexity in its process and function. In T.M. Bey & C.T. Holmes (Eds.), *Mentoring: Contemporary principles and issues* (pp. 79-93). Reston, VA: Association of Teacher Educators.

Heckman, P. (1987). Understanding school culture. In J. Goodlad (Ed.), *The ecology of school renewal, 86th yearbook for the National Society for the Study of Education* (pp. 63-78). Chicago: NSSE.

Hetlinger, J.S. (1986). Teacher orientation: What principals should stress. *NASSP Bulletin, 70*(491), 92-93.

Huling-Austin, L. (1989). Beginning teacher assistance programs: An overview. In L. Huling-Austin, S.J. Odell, P. Ishler, R.S. Kay, & R.A. Edelfelt (Eds.), *Assisting the beginning teacher* (pp. 5-18). Reston, VA: Association of Teacher Educators.

Huling-Austin, L. (1990). Teacher induction programs and internships. In W.R. Houston (Ed.), *Handbook for research on teacher education*. New York: Macmillan.

Huling-Austin, L., Odell, S.J., Ishler, P., Kay, R.S., & Edelfelt, R.A. (Eds.). (1989). *Assisting the beginning teacher*. Reston, VA: Association of Teacher Educators.

Hunt, D. (1968, October). Teacher induction: An opportunity and a responsibility. *NASSP Bulletin, 52*(330), 130-135.

Hunt, D. & Associates (1968, May). Teacher induction: A key to excellence. *NASSP Bulletin*, 68-78.

Johnson, S.M. (1990). *Teachers at work: Achieving success in our schools*. New York: Basic Books.

Kilcher, A. (1989). *Mentoring beginning teachers: A review of theory and practice*. Unpublished paper, The Learning Consortium, Toronto, ON.

King, A.J.C., & Peart, M.J. (1992). *Teachers in Canada: Their work and quality of life*. Ottawa, ON: Canadian Teachers' Federation.

Lieberman, A. (Ed.). (1986). *Rethinking school improvement: Research, craft, and concept*. New York: Teachers College Press.

Leithwood, K. (1990). The principal's role in teacher development. In B. Joyce (Ed.), *Changing school culture through staff development* (pp. 71-90). Alexandria, VA: Association for Supervision and Curriculum Development.

Leithwood, K.A. (1992). The principal's role in teacher development. In M. Fullan & A. Hargreaves (Eds.), *Teacher development and educational change* (pp. 86-103). London, UK: Falmer Press.

Little, J.W. (1987). Teachers as colleagues. In V. Richardson-Koehler (Ed.), *Educator's handbook: A research perspective* (pp. 491-518). New York: Longman.

Little, J.W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91*, 509-536.

Little, J.W., & Nelson, L. (Eds.). (1990). *A leader's guide to mentor training*. San Francisco: Far West Laboratory for Educational Research and Development/ERIC Clearinghouse on Educational Management.

McEvoy, B., & Morehead, M.A. (1987). Teacher induction: What can a university do best? *Action in Teacher Education, 8*(4), 45-49.

Merriam, S.B. (1988). *Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.

Miles, M.B. (1990). New methods for qualitative data collection and analysis: Vignettes and prestructured cases. *Qualitative Studies in Education, 3*(1), 37-51.

Neal, J.C. (1992). Mentoring: A teacher development activity that avoids formal evaluation of the protégé. In T.M. Bey & C.T. Holmes (Eds.), *Mentoring: Contemporary principles and issues* (pp. 35-49). Reston, VA: Association of Teacher Educators.

New Brunswick Commission on Excellence in Education. (1992). *Schools for a new century*. Report to the Commission on Excellence in Education. Fredericton, NB: Author.

Odell, S.J. (1989). Developing support programs for beginning teachers. In L. Huling-Austin, S.J. Odell, P. Ishler, R.S. Kay, & R.A. Edelfelt (Eds.), *Assisting the beginning teacher* (pp. 19-38). Reston, VA: Association of Teacher Educators.

Odell, S.J. (1990). Support for new teachers. In T.M. Bey & C.T. Holmes (Eds.), *Mentoring: Developing successful new teachers* (pp. 3-23). Reston, VA: Association of Teacher Educators.

Rosenholtz, S.J. (1987). Workplace conditions of teacher quality and commitment: Implications for the design of teacher induction programs. In G.A. Griffin & S. Millies (Eds.), *The first years of teaching: Background papers and a proposal* (pp. 15-34). Chicago, IL: University of Illinois at Chicago.

Rosenholtz, S.J. (1989). *Teachers' workplace*. New York: Longman.

Sarason, S. (1971). *The culture of the school and the problem of change*. Boston, MA: Allyn and Bacon.

Sarason, S. (1982). *The culture of the school and the problem of change* (2nd ed.) Boston, MA: Allyn and Bacon.

Schlecty, P., & Vance, V. (1983). Recruitment, selection, and retention: The shape of the teaching force. *Elementary School Journal, 83*, 469-487.

Scriven, M. (1988). Evaluating teachers as professionals: The duties-based approach. In S.J. Stanley & W.J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 110-142). Alexandria, VA: Association for Supervision and Curriculum.

Shulman, J.H., & Colbert, J.A. (Eds.). (1987). *The mentor teacher casebook*. CA: Far West Laboratory for Educational Research and Development, ERIC Clearinghouse on Educational Management, and ERIC Clearinghouse on Teacher Education.

Thiessen, D., & Kilcher, A. (1991). *A review of recent literature on innovations in teacher education*. Report to the Ontario Ministry of Education, Toronto, ON.

Tuinman, J., & Brayne, R. (1988). *The on-site personnel who facilitate learning: Commissioned papers: Volume 4*. Report to the British Columbia Royal Commission on Education, Vancouver, BC.

Varah, L.J., Theune, W.S., & Parker, L. (1986). Beginning teachers: Sink or swim? *Journal of Teacher Education, 37*, 30-34.

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*, 143-178.

Wideen, M.F., & Andrews, I. (Eds.). (1987). *Staff development for school improvement*. Philadelphia, PA: Falmer Press.

Wildman, T.M., Magliaro, S.G., Niles, R.A., & Niles, J.A. (1992). Teacher mentoring: An analysis of roles, activities, and conditions. *Journal of Teacher Education, 43*(3), 205-213.

Zeichner, K., & Gore, J. (1990). Teacher socialization. In W.R. Houston (Ed.), *Handbook of research on teacher education* (pp. 329-348). New York: Macmillan.

Zeichner, K., & Tabachnick, B.R. (1985). The development of teacher perspectives: Social strategies and institutional control in the socialization of beginning teachers. *Journal of Education for Teaching, 11*, 1-25.

Zimpher, N.L., & Rieger, S.R. (1988). Mentoring teachers: What are the issues? *Theory Into Practice, 27*, 175-182.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 319-336*

Paul T. Begley
Ontario Institute for Studies in Education

Contemporary Images of School Leadership in Ontario, the Northwest Territories, and Western Australia

This article reports three university-sponsored action research projects carried out during 1990-1992 in the Province of Ontario, the Northwest Territories, and Western Australia. The projects involved the development of multilevel, situationally sensitive profiles of contemporary school leadership practices by writing teams composed of 12 to 15 practicing school administrators and one or two university personnel. The outcome was three documents, each depicting in a graphical manner contemporary and regionally specific images of the principalship. Each profile portrays multidimensional images of practice, described behaviorally within developmental stages of growth leading to ideal practice as locally defined. Other outcomes include the opportunity to compare and contrast regional images of the principalship, the intrinsic value of the experience reported by the members of the writing teams, and several useful applications of the profiles as professional development resources at the provincial, district, and school site levels.

Cet article fait le reportage des trois projets de recherche-action dans la province de l'Ontario, dans les Territoires du Nord-Ouest, et dans Australie-Occidentale commandités par différentes universités. Dans le cadre de ces projets, on cherchait à poursuivre le développement de divers niveaux de profiles délicats in situ de la pratique administrative scolaire contemporaine par des équipes d'écrivain(e)s composées de 12 à 15 administrateurs et administratrices scolaires pratiquant(e)s ainsi qu'un ou deux représentants du personnel universitaire. Le résultat se présentait sous la forme de trois documents distincts, chacun illustrant de manière graphique un portrait spécifique du principalat régional contemporain. Chaque profile démontre le portrait à plusieurs dimensions de la pratique administrative décrite comportementalement selon les stages développementales de croissance menant à la pratique idéale telle qu'identifiée localement. On peut inclure parmi les autres résultats issus des projets ceux: d'avoir l'opportunité à faire la comparaison et le contraste entres les images régionales du principalat; de la valeur intrinsèque à l'expérience de la rédaction des textes par les écrivain(e)s; et ceux des différentes applications possibles de ces profiles comme ressource de développement professionnelle au niveau provincial, au sein des districts scolaires et sur place dans les écoles-mêmes.

Several persistent problems confound the efforts of those committed to the development of expert school leadership practices in Canada and Australia. One challenge has been the sheer number of available images of school leadership that have emerged from the literature in the last decade alone (Leithwood, Begley, & Cousins, 1990). As a result, consensus on what constitutes the nature of ideal school leadership and how it should be defined has been difficult to establish among academics and practitioners alike. A second problem derives from the dynamic nature of the school leadership role itself. When the scope of

Paul Begley is an associate professor in the Department of Educational Administration. His areas of academic interest include administrative values, school improvement processes, and leadership development in cross-cultural environments.

inquiry extends beyond the usual four- or five-year span of concern, it is apparent that expectations associated with the role of principal have been in a continual state of flux in response to a parade of situational influences and social trends. A third difficulty emerges from the rearview mirror perspective of most formal preparation programs and the questionable pedagogy and relevance of many university and field-based professional development projects. Finally, it is our observation that most formal preparation programs fail to accommodate divergent regional needs and the varying learning readiness levels evident among the participants in such programs. Can we reasonably expect the prescriptions for effective school leadership derived from American research to have equal relevance to school administration in Australia or the Canadian Arctic? What are the differences, if any, in the expectations associated with school leadership in these widely separated regions?

Our efforts to respond to these challenges led us to explore the development and use of situationally specific, multidimensional profiles of professional practice as a basis for fostering the development of expert school leadership practices. These contemporary profiles of school leadership are grounded in locally defined images of effectiveness as well as the findings of recent research in this field. As an outcome of these field development projects, leadership profiles are now being employed as a basis for designing preservice and inservice professional development programs in three relatively far-flung regions of the world; Western Australia, Ontario, and the Northwest Territories. Also, practitioners in these three regions now have an alternative to the traditional formal training program in the form of leadership profiles that can be used by individuals or small groups for self-directed and/or formative professional development.

All three of the regional profiles of effective school leadership share a common structure and development process. They take the form of two dimensional matrices that describe stages of professional action within selected dimensions of professional practice. The profiles were developed by writing teams made up of academics and practitioners working collaboratively in given regions in an action research mode. The choice of location for these projects was largely precipitated by our established presence as sponsors of formal preservice and inservice principal preparation programs in those regions. Despite the apparent diversity of the locations selected for the profile development projects, the school administrators of all three regions were found to be confronting a common challenge, the need to change their approaches to school leadership in response to a range of social and educational pressures. Of course, the specific nature of the social or educational pressures varied from region to region. The common thread, however, was the perception that traditional role orientations were no longer adequate for the demands of the job.

In addition to the utility of the documents themselves, and the value to the members of the writing teams of the time spent deliberating on professional practices, the development of these profiles has provided us with the unique opportunity to compare and contrast the images and dimensions of effective school leadership practices in three distinct regions. I begin this article with a

brief review of some generic problems associated with school leadership development. The discussion then turns to the conceptual underpinnings and procedures of the profile development process. I include a discussion of several issues and liabilities that have emerged from our work. This is followed by a description and comparison of the three profiles that have been developed and some discussion of how they are currently being used to promote effective school leadership.

What the Research Says: Alternate Conceptions of the Principal's Role

Existing research has conceptualized principal practices in a variety of ways; as unidimensional (Cuban, 1986) or multidimensional (Sergiovanni, 1984), in terms of typical practices (Kingdon, 1985) and effective practices (Larsen, 1987), or as actual versus preferred roles (Gousha, 1986). Other researchers have preferred to conceptualize the role by describing patterns of practice or leadership styles (Hall, Rutherford, Hord, & Huling, 1986 or Leithwood & Montgomery, 1986). More recently a cognitive perspective on the school leadership role has come into vogue that focuses on the internal and external influences on the principalship (Leithwood & Hallinger, 1993; Leithwood, Begley, & Cousins, 1992).[1]

Except for the latter, these research conceptualizations of the principalship have tended to be descriptive or correlational. They have not been particularly informative on the intents of administrative actions or the underlying and motivating values of the actors. Often schools and principals may be described as effective without saying much about how they become effective. The focus has been on the empirical and technical, and the bulk of the literature has been blind to the more moral aspects of leadership, at least until recently.

From the perspective of many practitioners, such research is too abstract, too lacking in the relevance, utility, and cultural specificity that would provide useful guides to practice. Research findings may even promote artificial perceptions of uniformity among school organizations and false images of homogeneity within school leadership roles. Most prescriptions for expert school leadership simply do not ring true for administrators who must work in specific contexts. Yet we propose that these same research findings remain our best available resources when it comes time to train, develop, or promote school leadership practices. In effect we have the content, but we need a process that will build relevance and utility for individuals leading schools in particular settings.

The positive aspect of this situation is that principal preparation programs can now be grounded in more than just the context-bound practices, or theories-in-use (Argyris, 1982), manifested by local practitioners serving as instructors. This availability of research-derived knowledge on effective administrative practices has permitted the design and development of much better validated leadership programs for principals (Daresh & Playko, 1992). However, despite the increased rigor evident in research-driven program designs, several issues remain. Most relate to candidate perceived relevance as alluded to above. Among practitioners the relevance of research derived knowledge is largely assessed on the basis of how easily such information can be applied in specific contexts and translated into administrative action. In our experience several factors can influence the candidate perceived relevance of

leadership development programs. In the following section of this article several of these factors are identified and discussed.

School Leadership Development: Practical and Conceptual Issues

The design and implementation of formal training programs that meet the current and future needs of school leaders should involve considerably more than just mirroring the typical practices of the past. Effective programs will be those that address the knowledge, skills, and attitudes required by the leaders of future schools. Furthermore, such learning and skill acquisition must be achieved through a process or learning context that is responsive to several factors. An ideal learning environment will be sensitive to the varying levels of readiness manifested by school leadership candidates on entry into a formal preparation program. It will also recognize the particular needs of adult learners. Finally, the delivery of even an exemplary program will inevitably be complicated by the precedents and expectations established by previous leadership programs and even the potentially divergent orientations of instructional personnel and program sponsors.

Probably the first question to consider is which image of leadership one should adopt as an ideal. In Ontario, the Northwest Territories, and Western Australia virtually all principal preservice preparation programs now manifest varying degrees of commitment to the view of the principal's role usually described as instructional leadership.[2] It is possible to argue convincingly in support of more recently developed conceptions of the role, such as transformational leadership. However, given the mandated nature of the syllabus for most principal certification programs, and the natural time lag encountered when new knowledge impacts on established procedures, we initially concluded that instructional leadership remains the dominant image of school leadership. In execution, that proved not necessarily to be the case for at least two of the profiles that we developed. With the Northwest Territories Profile, the writing team was keen to incorporate in their document what they viewed as the highly attractive, minority group empowering vocabulary associated with the literature of transformational leadership (Burns, 1978; Leithwood, 1992). An earlier source, Miller and Seller (1985), identifies the transformational orientation as one aimed at achieving personal integration and particularly social awareness. The strategies of the transformation orientation are those of creative thinking, invitational teaching, cooperative learning, guided imagery techniques, and whole language learning. A similar attraction to the concepts of transformational leadership was noted with the Ontario profile writing team, although they chose to avoid the term *transformational* as much as possible because of its perceived trendiness.

Given the array of educational experiences and skills required for the development of a comprehensive image of leadership effectiveness, it also seems important to take a longitudinal and developmental view of administrators' socialization experiences. Developmental principal profiles such as the three described here at least potentially could play an important role in supporting this theorized process of professional maturation; particularly at what van Gennep (1960) terms the transition and incorporation stages of development. We find van Gennep's three stage model of professional socialization (separation, transition, and incorporation) helpful in conceptualizing the develop-

mental process whereby individuals are influenced by several socializing factors (Ronkowski & Iannaccone, 1989). Van Gennep (1960) suggests a process of general professional maturation progressing from being defined by others to being self-defined. According to van Gennep, at the separation stage people are concerned with comparing themselves with others and how others judge their adequacy. At the transition stage the pattern of *self-location* is against the standards imposed by the functions of the job and task performance. At the incorporation stage individuals make comparisons between their former and present self (perceptions of progress made from a previous self toward becoming an instructional leader).

As any experienced staff developer can attest, semantic problems occur with the use of particular educational or academic jargon, more or less acceptable from region to region. From a more academic perspective, we know that anthropologists and sociolinguists regularly analyze language to derive insights into a culture's roles, norms, taboos, values, and world views (Spradley, 1979 cited in Marshall, 1992). More to the point, in our experience a particular emphasis on some dimension of practice in one region can easily alienate the audience from another region. For example, discussing *socialization* research with a group of practitioners in Australia may suggest mind control, a negative connotation, whereas with a Canadian audience, not a ripple would be noted. Similarly, in many regions of Canada the term *child-centered education* has recently become politically charged and is to be avoided at all costs. On the other hand, American practitioners seem much inclined to speak of *test scores* and improving a school's *grade point average*, terms that are not user-friendly in the Australian and Canadian interface. The point is that the development and use of research-driven but situationally sensitive profiles may prove helpful in addressing this issue. Thus a profile for an Australian audience might safely address the *pastoral* responsibilities of the principal without fear of confusing Canadian principals who likely would associate this term with crop management or animal husbandry.

A recent study on the socializing influences experienced by aspiring principals (Begley & Campbell-Evans, 1993) lends further support to the notion that developmental profiles may be useful to school leadership development. According to the findings of this study involving 87 aspiring principals in the Northwest Territories, more than half the aspirants apparently pursued principal training as a way of moving toward an existing image of the principal's role. More than 50% also cited skill development as motivating them to pursue training and nearly a third of the respondents sought training to broaden their professional perspective. The implication for those interested in promoting principal recruitment, or more to the point encouraging the professional development of individuals, is that the best strategy may be to emphasize the organizational *career ladder* and *skill development* aspects of principal training, and the general value and applicability of the training to presently held positions. This is consistent with Van Maanen's notion of *chains of socialization* (Feldman, 1989), which suggests that the process of organizational socialization consists of aspirants making only minor modifications in their behavior and attitudes from job setting to job setting.

A final word before we address the methodology of profile development. We wish to acknowledge the perspective of radical theorists, subjectivists, and the like. Our orientation to profile development has been, without apology, functionalist. To produce our profiles we employ a set of procedures derived from and founded on traditional curriculum gap analysis techniques (Leithwood & Montgomery, 1987), planned educational change theory (Fullan, 1991), and the procedures of school improvement (Huberman & Miles, 1984; Leithwood, Fullan, & Heald-Taylor, 1987). Thus our work reflects an interventionist orientation toward the practice of school leadership. On the other hand, we do acknowledge the contingent and dynamic nature of the school leadership role, as well as the multiple images of situationally and subjectively defined practice, hence our interest in multiple, regularly validated images of the role. Furthermore, the members of the writing team were for the most part practicing principals and therefore the primary stakeholders in the process of profiling their own practices. However, these are people who do not question their right or responsibility to exercise leadership in their schools. Finally, although these profiles reflect the findings and vocabulary of school improvement and principal effectiveness research, they are primarily grounded in the actual practices of keen and competent school leaders who are much valued as professionals in their own jurisdictions.

What is a Profile?

A profile is a two-dimensional matrix that describes developmental stages of growth in professional action in selected dimensions of professional practice. The creation of a profile begins with the establishment of a goal statement followed by a series of decisions about which categories of professional action are most relevant to the achievement of the desired state described in the profile goal statement. In a profile these categories are called dimensions. Each of these dimensions is also usually broken down into a set of subdimensions. To accomplish this, various facilitative and consensus building strategies are employed to blend research findings from the literature with local craft know-

1. *Advocacy:* The principal develops a shared vision which supports the educational needs of students and the aspirations of their community.
 - Community Values
 - Empowerment
 - Community Education Plan
 - Student and Staff Support
2. *School Culture Management:* The principal creates and maintains a supportive school climate which is conducive to learning.
 - School Environment
 - Decision Making
 - Language and Culture Promotion
3. *Instructional Leadership:* The principal initiates and directs a growth-oriented process to maximize learning outcomes for staff, students, and community.
 - Planning
 - Development and Implementation
 - Evaluation
4. *Organizational Management:* The principal ensures the effective operation of the school.
 - Finances
 - Facilities
 - Human Resources
 - Policies and Procedures
 - Time Management

Figure 1. Key dimensions and subdimensions of principal practice (Northwest Territories).

The Principal as Visionary: Recognizing a broad range of environmental influences, is capable of perceiving a condition beyond that which is immediately apparent and of employing that view to provide focus and direction to the activity occurring within the principal's sphere of influence.

Vision Development	possesses selected or isolated ideas and views which may become the components of a vision	→ formulates, and articulates for the teaching staff, a vision for the school	→ collaborates with the teaching staff to articulate a school vision which provides purpose and direction to school affairs	→ collaborates with representative members of the school community to articulate a vision which provides purpose and direction to school affairs
Vision Derived Goals	possesses a set of goals derived from Ministry and Board expectations	→ develops school goals consistent with the principal's articulated vision	→ works with the teaching staff to develop school goals which reflect their collaborative vision	→ collaborates with representative members of the school community to develop goals which reflect the community developed vision statement
Communication of the Vision	shares selected ideas and views with the teaching staff	→ shares a vision for the school with the staff	→ interacts with the teaching staff to obtain feedback which may serve to redefine, redirect or prioritize school initiatives	→ communicates regularly with the community, by various means; sharing information and obtaining feedback which may serve to redefine, redirect or prioritize school initiatives
Change Agent	senses a need for change within the school	→ initiates changes based on the principal's articulated vision	→ with teaching staff, develops and implements appropriate strategies to effect such change as may be required to realize school goals	→ collaborates with staff to develop and implement strategies aimed at achieving school and community goals; with particular attention to the ongoing professional development of staff
Culture	responds to traditional cultural norms such as orderliness, academic achievement and co-curricular participation	→ works to establish a school culture which supports the principal's vision for the school	→ encourages teaching staff to participate in formulation or refinement of a school culture consistent with the school vision	→ establishes / maintains a collaborative school culture which encourages all school partners to be responsive, creative & visionary

Figure 2. Alternate levels of practice from the Ontario principal profile dimension entitled Principal as Visionary.

ledge. For example, in the NWT Principal Profile, the writing team ultimately settled on four key dimensions of principal practice: Advocacy, School Culture Management, Instructional Leadership, and Organizational Management (Figure 1).

The next step in creating a profile is behaviorally describing, in each dimension or subdimension, the range of professional behavior that might be observed in the work setting. These alternate levels of professional practice are sequenced in dimensions of the profile according to their relative impact on attaining the desired outcomes identified in the profile goal statement. Figure 2, taken from the Ontario profile, illustrates this. The levels can be thought of as the stages of growth ranging from typical competent practice to highly exemplary or ideal practice. Profiles are generally not employed to describe incompetent practices.

Profiles are developed by a writing team of representative practitioners and academics who generate a series of draft documents. This approach is consistent with the tenets of action research that propose a mode of interpretive contextual inquiry for the purposes of understanding and describing events as lived, and for directly informing practice and its embedded and related theories (Cole & Knowles, 1993). Draft versions of the profiles produced by these writing teams typically undergo an extensive field validation process before a final version of the document is released for use. Developing each of the profiles described in this article typically required 30 to 40 hours of intensive collaborative group work by the actual writing team, plus many additional days of work by a subset of the team in editing and validating the various drafts. Full validation of the final drafts has typically required at least one academic year. In our view, principals' role profiles have a relatively short shelf life because of the rapid rate of change associated with school leadership practices. They should probably be reviewed and revalidated within five years.

The original and probably best known principal profile was produced in 1983-1984 by Leithwood and Montgomery (1986). This was a rigorous application of the innovation profiling technique, essentially a curriculum analysis procedure, to a role description task (Leithwood & Montgomery, 1987). This profile was based on an exhaustive review of existing research on effective schools and instructional leadership practices. Following usual academic protocol, field validation with a large sample of practitioners occurred after the literature review. We reversed this procedure in developing our three profiles to the extent that we employed personal inventory surveys and consensus building activities with the writing teams before introducing and reviewing research findings. This reflects our more subjective philosophical orientations, as well as our desire to ensure a high degree of local relevance in terms of vocabulary and image definition.

Two attributes were common to all three of the profiling projects reported in this article; the time frame and the methodology employed. All three profiles were produced during 1990-1992 with myself acting as group facilitator for all three writing teams. Second, except as noted above, the procedure used to produce these developmental images of the principalship was based on an innovation profile development process proposed by Leithwood and Montgomery (1987). Applied as a role analysis procedure, this process

produces multidimensional images of practice described behaviorally in developmental stages of growth leading from typical competent practice to an image of ideal practice.

Using Principal Profiles:

Role profiles can be used for a variety of purposes. In addition to their obvious utility as a research-driven, field-validated conceptual framework from which to design and develop preservice and inservice programs, they can also be used for a number of less formal but highly desirable professional development purposes. School principals, working alone or with one or more partners, can use principal profiles as a resource to monitor and support their continued professional growth as school leaders in their jurisdictions. This may occur in several specific ways:

1. A profile may help to focus the principal's goals on the needs of students.
2. A profile may identify relevant criteria and standards of practice for self-evaluation, mentoring, or coaching purposes.
3. A profile integrates the findings of recent research on principal effectiveness and local craft knowledge in detailed and situationally relevant descriptions of principal action.
4. A profile may help individuals analyze the amount of change required in their professional practices into incremental steps leading toward some ideal image of the role.
5. A profile may help principals to identify, emphasize, and justify, or conversely to challenge, particular managerial and leadership practices identified by research and their colleagues as desirable.
6. A profile may provide a framework for predicting the obstacles that may be encountered by principals as they implement changes in their practices.
7. A profile may provide a basis for identifying school and regional differences that justify variations in an individual principal's practices.

All three profiles have now been published[3] as professional development resource documents and have been broadly distributed in their regions. They are primarily being used by individuals for personal professional development or adopted by school districts as an espoused image of effective school leadership practice. However, as alluded to above, the profiles are also being used for more formal purposes. As of 1992, the NWT Principal Profile has been adopted as the conceptual framework for developing the annually delivered and legislatively mandated principal preparation program, replacing Leithwood and Montgomery's (1986) profile of the principalship. As of 1993 the same has occurred in Ontario for all principal certification courses sponsored by the Ontario Institute for Studies in Education. In Western Australia the Primary Principals' Association has broadly inserviced their WA Principal Profile. In both Ontario and Western Australia the profiles have also influenced the design and development of professional development experiences sponsored at both the university and school district level.

Some Liabilities Relating to Profile Use

We can identify a number of potential liabilities resulting from the use of profiles. Although we believe that ultimate responsibility for appropriate and ethical use of such resources will lie with the user, this does not eliminate our

responsibility to consider and identify potential liabilities. We can readily identify these six limitations to profile use:

1. Profiles are expensive to produce in terms of the time and resources needed. Not all jurisdictions can afford to produce such labor intensive resource documents. This may raise an equity issue. In an era of fiscal restraint, some groups may devote less energy to field-validation than would be considered ideal, or employ something less than a representative group as a writing team.
2. Once a profile has been produced, it needs to be properly implemented. Project sponsors should avoid expectations for a quick fix in terms of increased principal effectiveness. The writing team members must also have credibility with the potential users of the profile; otherwise implementation efforts will fail. Finally, a profile should be attractively laid out in order to invite use by busy school administrators who already must process a lot of paper.
3. Most profiles are quite linear and sequential in their portrayal of a role. They are also more often than not complex documents. This may offend individuals who are more intuitive in their orientations. Moreover, a profile is in many respects a caricature of professional practice, emphasizing some dimensions and not others. Some individuals whose professional practices are not reflected in the dimensions and developmental stages may become alienated. Nobody likes to be perceived as incompetent or out of step professionally.
4. Some educators are strongly inclined to be innovation focused. They may mistakenly view the profile as an *end* in itself rather than a *means* to an end. Dogmatic adherence to a fixed role image is not a desirable outcome of profile development. The nature of the principalship is in continual flux and this dynamism in response to social and environmental influence will likely continue.
5. There is a very real potential for the misuse of principal profiles as a tool of summative evaluation. Most principals involved in the validation of our profiles have been quick to point this out. In our view this is not an appropriate use for profiles. Informal and formative or growth-oriented performance appraisal applications may be acceptable. However, there are simply too many personal and contextual variables that cannot be accounted for to allow the use of profiles for summative evaluation purposes. We acknowledge that there are those who do not agree with us.
6. Profiles must be reviewed and updated from time to time. Changing role requirements, new research findings, a continually evolving image of what it is to be an educated person are all factors that conspire to give profiles a relatively short shelf life; probably five years or less. They ought to be reviewed and renewed on a cyclical basis just like school curriculum.

The Three Profiles

In this section the results of the profiling projects, three separate regional profiles of the principalship, are described. The profiles were produced by writing teams in the Northwest Territories (NWT), Ontario, and Western Australia.

The NWT Principal Profile: Instructional Leadership and Community Facilitation

The majority of schools in the Northwest Territories (a vast region of the Canadian Arctic located north of the 60th parallel) are situated in small, isolated communities. The population of the Northwest Territories (NWT) is predominantly indigenous (Inuit, Dene, and Métis) except in Yellowknife, the capital. Despite these unique features of context, a needs assessment process conducted by the Ontario Institute for Studies in Education for the NWT Department of Education in 1987 identified instructional leadership, a popular conception of school leadership in Canada and the USA, as equally relevant to school administration in Arctic Canada. This was perhaps not as surprising as we might have been inclined to think at the time—all but a few of the incumbent school principals were non-Native, much of the formal school curriculum has been adopted from the southern regions of Canada, and most of the teachers employed in the north acquire their professional training at colleges in the south.

Since 1987, however, academics and practitioners associated with the NWT legislated principal certification program have increasingly recognized the shortcomings of instructional leadership as an image of ideal practice on which to base their program. It may be an appropriate role image for large, white-dominated communities such as Yellowknife, but it has clearly failed to address important expectations for the role in small, isolated communities. Other factors of concern were the persistently high turnover rates for principals and the continuing low levels of recruitment to the principalship from among Aboriginal educators.

The publication in 1991 by the NWT Department of Education of a new document entitled *Our Students, Our Future: A Planning Framework* produced the final impetus necessary for launching a project to develop an updated profile of the school leadership role in the NWT. This document identifies three areas of responsibility for principals in the NWT, at least one of which extends beyond the normal scope of instructional leadership. The first two, responsibility for instructional programs and creating a supportive learning environment were familiar enough. However, the third, provision of services that facilitate students' physical and cognitive preparedness for learning, implies for the NWT school administrator a broader, more community-wide scope of concern.

The profiling project was sponsored by the NWT department of Education and the Ontario Institute for Studies in Education. A writing team was organized, composed of four practicing school principals, three district administrators, three Department personnel, one OISE faculty, and an OISE research officer. They met for four consecutive days to produce the first draft of the profile in May 1992. Two of the team members were Aboriginal practitioners and four were female educators. Early on in the process they established their goal for the NWT Principal Profile "to identify key dimensions of practice through which principals meet the needs of individual students, improve the quality of teaching and learning, and support the aspirations of the community." Without abandoning the now familiar core dimensions of instructional leadership (collaborative goal-setting, direct participation, monitoring, and a concern for student outcomes), the team produced a more situationally

sensitive image of the role that reflects the cultural base of the Northwest Territories.

Several new dimensions or subdimensions of the principal's role were proposed including "consensual team building," "inter-agency community partnerships," and "community facilitation." However, the four key dimensions of principal action that emerged in the final version of the profile were advocacy, school culture management, instructional leadership, and organizational management. Final validation of the interim profile was carried out during 1993 and it was published as a resource document in June 1993.

The Ontario Profile: Visionary Principals and Problem Solving Processes

Several factors combined to produce a school leadership profiling initiative in Ontario. One important motivation was the Ontario Ministry of Education. During 1990 the Principals' Course Advisory Committee of the Ontario Ministry of Education articulated a need to review and update the core objectives for its certification program. The existing core objectives had been developed in 1984 and focused primarily on the curriculum management role of the principal. By 1990 it had become increasingly apparent that other important expectations for the role were not being addressed by the mandated principal certification program. Several new notions needed to be incorporated such as transformational leadership, collaborative school cultures, and community empowerment.

During the same time frame, the senior administrators of a school district located in southwestern Ontario expressed to OISE faculty an interest in examining the links between the principal's role in district level strategic planning and school-based improvement projects. They had committed a considerable amount of district energy toward a top-down strategic planning initiative that produced detailed objectives for school improvement. Unfortunately, the links between the district plans and what individual principals were expected to do in schools to support these plans was less than clear. It was hypothesized that a regional principal profile might help bridge this gap in the planning process.

A final source of energy for the Ontario Profile initiative came from a third school district, in this case a publicly funded Roman Catholic school system. Much like the Western Australian principals discussed below, these Ontario principals expressed a desire to use the innovation profile procedure as a needs assessment process before developing inservice programs that would link district priorities and initiatives to their personal professional development plans.

A writing team of 15 individuals was established with representation from five school districts. A faculty member from OISE, aided by a research officer, acted as coordinators and facilitators for the process. Seven members of this writing team were female. They met for a total of six days spread across several months in early 1992. The team identified as their profiling goal "the identification of the key dimensions of professional practice for principals committed to improving the quality of education." After several days of deliberation, five key dimensions were selected to describe the principalship in Southwestern Ontario. These are: the principal as visionary, the principal as problem solver,

the principal as instructional leader/program facilitator, the principal as school community facilitator, and the principal as manager.

A Principal's Profile (Western Australia, 1990)

Recent changes in the Western Australian education system resulting from the release and implementation of a state policy document entitled *Better Schools in Western Australia: A Programme for Improvement* (Ministry of Education in Western Australia, 1987), induced significant changes in the nature of the Western Australian primary principalship. Principals were now required to play a much expanded leadership role in their school communities compared with the past—augmenting traditional management functions with relatively new expectations for instructional leadership. Partially in response to these pressures, a project was launched to produce a profile of the primary principalship in Western Australia (WA). This profile, developed in 1990, was the first of the series of three principals profiles that we developed. The resulting document was primarily intended to be used by incumbent principals as a resource to promote their own self-development and empowerment. This is in contrast to the agenda common to the Northwest Territories and Ontario projects aimed at producing profiles to support the development of formal professional development programs.

The WA Principal's Profile was developed by a writing team made up of principals from the Ministry of Education and the Catholic education system. As is the case for the other two profiles developed later, its development was largely based on the collective experiences of the writing team as well as a literature review of key Ministry and research documents. The stated objective of the profile was, "to demonstrate collaborative leadership that will optimize the education of each member of the school community." As is the case for the other profiles, it was not intended as a deficit model that describes the shortcomings of a failing principal. The user of the profile is assumed to be performing at a level considered at least satisfactory.

The WA Principal's Profile identifies three key dimensions: self, which describes the personal qualities of successful principals; processes, which describe gradations of successful use of processes; and outcomes, which describe the goals of the successful principal's actions.

Discussion and Outcomes of the Profiling Projects

In this section I compare the three profiles and discuss the outcomes and applications that have resulted from the projects.

The profile dimensions. When the three profiles are compared it becomes evident that all share a common core of three key dimensions (see Figure 3). With semantic differences, all three profiles have in common descriptors of effective practice relating to school culture management, instructional leadership, and organizational management. In some ways this is a remarkable outcome given the far-flung distances and variations in culture involved. It suggests that there are important similarities in the Australian and Canadian principalships. In both countries principals apparently see their contemporary roles as extending much beyond traditional building management to include instructional and cultural leadership responsibilities. On the other hand, given that all three profile teams reviewed the same school improvement research

Key Dimensions of Principal Practice: *Western Australia*	*Ontario*	*Northwest Territories*
1. Self	*1. Visionary*	*1. Advocacy*
- Vision	- Vision Development	- Community Values
- Problem Solver	- Vision Derived Goals	- Empowerment
- Effective Communicator	- Communicator of the Vision	- Community Education Plan
- Change Agent	- Change Agent	- Student & Staff Support
- Monitor	- Culture	
2. Processes	*2. Instructional Leader*	*2. Instructional Leadership*
- *Instructional*	- Learning Environment	- Planning
- Goal Setter	- Goals: Sources and Uses	- Development and Implementation
- Curriculum Leader	- Leadership Strategies	- Evaluation
- Programme Evaluator	- Implementation	
- Expectation Holder	- Program and Staff Evaluation	
- *Pastoral/Cultural*	*3. School-Community Facilitator*	*3. School Culture Management*
- Cultural Leader	- Goal Setting	- School Environment
- Acknowledger of Student Achievement	- Interagency Liaison	- Decision Making
- Acknowledger of School Community Achievement	- Culture Management	- Language and Culture Promotion
- Community Facilitator	- Communication	
	- Entrepreneurial Skills	
- *Management/Administrative*	*4. Manager*	*4. Organizational Management*
- Staff Developer	- Student Conduct	- Finances
- Human Resource Manager	- Budget	- Facilities
- Communicator	- Supervision of Personnel	- Human Resources
- Time Manager	- Regulations and Policy	- Policies and Procedures
	- Time Management	- Time Management
3. Outcomes	*5. Problem Solver*	
- Instructional	- Problem Interpretation	
- Pastoral / Cultural	- Goals and Values	
- Management / Administrative	- Solution Processes	

Figure 3. Comparing the key dimensions of practice from three profiles; WA, Ontario, and NWT.

findings as part of the profiling process, it is not surprising that the profiles share some of the same vocabulary and concepts.

A comparison of the NWT and Ontario profiles reveals a heavier emphasis in the NWT profile on transformational (collaborative, teacher, and community empowering) strategies. The ideal NWT principal is portrayed as engaging in school culture or organizational *management* as opposed to being *the manager*; an emphasis on facilitation and consensus building rather than control. In contrast, the Ontario profile projects a more individual or charismatic leadership-oriented image of the role. The Ontario Profile also presents a more ambitious or sophisticated image of the principalship. This is illustrated by the inclusion of *visionary* and *problem solver* as additional key dimensions of practice. However, in both profiles the writing teams have clearly expanded their image of the principal's role beyond the limitations of traditional instructional leadership.

The WA Principal Profile also has some unique qualities. The Australian writing team chose to place the three dimensions common to all three profiles (culture, instruction, and administrative management) in a single composite dimension they called *processes*. They then created a separate dimension, titled *self*, as a way of describing the personal orientations that drive the activities or processes of the principalship. This effort to isolate the beliefs and intents that underlie the actions of the principalship is unique to the three profiles.

A second unique feature of the WA Profile is a third composite dimension identified as *outcomes*. This key dimension collates the highest levels of effective practice identified under processes into a summary of goal oriented practices. Like the Ontario and NWT profiles, the three sections of the WA profile are clearly interrelated, but unlike the other two profiles may be used separately or as a total picture. Thus the WA Profile user may look at his or her success on the outcome dimension of the profile and then validate this against the processes and self-profiles. Alternatively the principal can start at processes and work either way to the other dimensions.

The profile growth strands. The preceding section focuses on the dimensions of practice selected by the writing teams for their regional profiles. In this section we look at the developmental growth described in the profiles, moving from the lower levels of competent practices to the ideals described at the highest level of the profiles. Comparison and analysis of these descriptive statements reveal the implicit *growth strands* that the writing teams incorporated in their profiles. Some intriguing images of professional growth emerge. What follows is a sampling of the growth strands common to all three profiles.

From a tendency toward reactive responses to proactive responses. The *management* subdimensions in both the NWT and Ontario profiles (finances, time management) illustrate this growth strand. For example, at the lowest level of practice the principal might "devote available administrative time to immediate administrative tasks and daily occurrences," whereas at a more ideal level the principal might "develop and use a time management plan to focus the use of personal time, as well as staff resources, toward the achievement of short and long term goals."

From reliance on self, to mimicry of others, to attention on outcomes, to sensitivity to multiple environmental influences. This growth strand is evident particularly in the school and community culture management dimensions of all three profiles. What is illustrated as one proceeds along the stages of growth is an increased awareness and responsiveness to educational stakeholders and educational influences derived from the broader school community. At lower levels the principal might be "aware of parental, staff and student impact on the culture of the school." At an ideal level the principal is more likely to "collaborate with students, staff and community to create and promote a culture in which unique needs and cultural diversity are respected."

From rigid adherence to fixed procedures, to procedural flexibility, to philosophical or conceptual fidelity. The instructional leadership dimensions in all three profiles illustrate this growth strand in such subdimensions as evaluation. At lower levels of competence, a principal might "conform to Board policy and use established procedures to evaluate staff" or "participate in mandated program reviews." A skillful instructional leader would "monitor progress toward attainment of school goals through program and staff evaluation procedures" and "work with teachers to develop self-evaluation procedures and professional reflection."

From in-school focus, to inter-school focus, to school within the greater community focus. All three profiles incorporate this growth strand, particularly in the school and community culture management subdimensions. In the Northwest Territories a competent principal would be "sensitive to the need for incorporating local cultural activities in the school" and "invite the community to attend special school events." However, more ideal practice would be demonstrated by a principal who "employs the cultures of the community to enhance individual and collective self esteem, and to promote learning in a multilingual educational system."

From a limited repertoire to a broad repertoire of strategies: The highest levels of practice across several dimensions of the three profiles, but especially the subdimensions of instructional leadership, manifest this notion. At lower levels of effectiveness principals might "employ a limited number of broadly accepted strategies to manage instructional affairs (inform staff of P.D. opportunities)," whereas at more exemplary levels principals might "recognize and encourage the development of curricular expertise among teachers and resource staff" and "promote sharing among staff to encourage skill development and reflection about teaching practices (collaborative peer coaching)."

Conclusion

The three profiles described in this article have now been published as professional resource documents and broadly distributed in their regions. Although there are potential liabilities resulting from the use of profiles, we believe that the benefits far outweigh the liabilities. Certainly the practitioners involved in the validation of these profiles have responded with enthusiastic support for the documents. We have been encouraged enough by the outcomes of our first three profiles that tentative plans have been made to produce at least two additional profiles in the coming year: one in the United States and a third Canadian profile.

Leadership profiles are currently being used in a variety of ways by Canadian and Australian administrators. Individuals employ them as tools to support personal professional development through self-evaluation, mentoring, and coaching. They have been adopted by several school districts as supports to an espoused image of effective school leadership practice, or to design needs assessment instruments for staff development purposes. More formal applications include their adoption as the conceptual framework for structuring mandated principal certification courses in Ontario and the Northwest Territories. Finally, in both Ontario and Western Australia their respective profiles have significantly shaped the design and development of professional development programs sponsored at both the university and school district level.

The popularity and ready acceptance of these leadership profiles by practitioners in Canada and Australia raises a final point. The importance of university/school district collaboration as a support to the renewal of education in North America is a recurring theme in the literature. Unfortunately, there are few documented examples of successful university/school district collaboration. Rarer still are projects that can be easily and economically replicated by other organizations wishing to develop inter-sector collaboration in their regions. We would propose leadership profile development as an example of the latter. Clearly there is a need for more field development projects of this type, that is, collaborative action research involving academics and practitioners working together to improve educational practice.

Notes

1. For a more comprehensive review of research on the principalship, see Leithwood et al. (1990) or Leithwood et al. (1992).
2. Instructional leadership is usually characterized by the presence of a clearly articulated educational philosophy, extensive knowledge about effective educational practices, and a clear understanding of the policy environment framing the school's purposes and practices. This image of the role emphasizes leadership as well as management functions and extends considerably beyond the performance of routine work tasks, the usual domain of training research. In the research community, justification for an instructional leadership approach to school administration has been based largely on the premise that certain characteristic actions of principals, intended to encourage and support classroom practices linked by research to improved student outcomes, have a positive impact on student achievement. However, research by Heck, Larsen, and Marcoulides (1990) has also begun to validate the instructional leadership role of principals as a causal link to improved student outcomes.
3. Copies of the published profile documents are available by contacting the author at Department of Educational Administration, Ontario Institute for Studies in Education, 252 Bloor Street West, Toronto, ON M5S 1V6.

References

Argyris, C. (1982). *Reasoning, learning, and action: Individual and organizational.* San Francisco, CA: Jossey-Bass.

Begley, P.T., & Campbell-Evans, G.H. (1993, March). Socializing experiences of aspiring principals. *Alberta Journal of Educational Research, 38,* 285-299.

Burns, J.M. (1978). *Leadership.* New York: Harper Row.

Cole, A.L., & Knowles, J.G. (1993). Teacher development partnership research: A focus on methods and issues. *American Educational Research Journal, 30,* 473-496.

Cuban, L. (1986). Principaling: Images and rules. *Peabody Journal of Education, 63,* 107-119.

Daresh, J.C., & Playko, M.A. (1992). *The professional development of school administrators.* Toronto: Allyn and Bacon.

Feldman, D.C. (1989). Socialization, resocialization, and training: Reframing the research agenda. In I.L. Goldstein & Associates (Eds.), *Training and development in organizations* (pp. 376-416). San Francisco, CA: Jossey-Bass.

Fullan, M.G. (1991). *The new meaning of educational change*. Toronto: OISE Press.
Gousha, R.P. (1986). *The Indiana school principalship: The role of the Indiana principal as defined by the principal*. Bloomington, IN: Indiana University School of Education.
Hall, G., Rutherford, W.L., Hord, S.M., & Huling, L.L. (1986). Effects of three principal styles on school improvement. *Educational Leadership, 41*(5), 22-31.
Heck, R.H., Larsen, T.J., & Marcoulides, O.A. (1990). Instructional leadership and school achievement: validation of a causal model. *Educational Administration Quarterly, 26*(2), 94-125.
Kingdon, H.D. (1985). The role of the teaching elementary school principal. *Principal Issues, 4*, 20-23.
Huberman, M.A., & Miles, M.B. (1984). *Innovation up close: How school improvement works*. New York: Plenum Press.
Larsen, T.J. (1987, April). *Identification of instructional leadership behaviours and the impact of their implementation on academic achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.
Leithwood, K.A. (1992). The move toward transformational leadership. *Educational Leadership, 49*(5), 8-12.
Leithwood, K.A., Begley, P.T., & Cousins, J.B. (1990). Nature, causes and consequences of principals' practices: An agenda for future research. *Journal of Educational Administration, 28*(4), 5-31
Leithwood, K.A., Begley, P.T., & Cousins, J.B. (1992). *Developing expert leadership for future schools*. London: Falmer.
Leithwood, K.A., Fullan, M.F., & Heald-Taylor, G. (1987). *School level CRDI procedures to guide the school improvement process*. OISE mimeo.
Leithwood, K.A., Hallinger, P. (1993). Cognitive perspectives on educational administration. *Educational Administration Quarterly, 29*, 296-301.
Leithwood, K.A., & Montgomery, D.G. (1986). *Improving principal effectiveness: The principal profile*. Toronto: OISE Press.
Leithwood, K.A., & Montgomery, D.G. (1987). *Improving classroom practices—Using innovation profiles*. Toronto: OISE Press.
Marshall, C. (1992). School administrators' values: A focus on atypicals. *Educational Administration Quarterly, 28*, 368-386.
Miller, J., & Seller, W. (1985). *Curriculum: Perspectives and practice*. New York: Longman.
Ministry of Education in Western Australia. (1987). *Better schools in Western Australia: A programme for improvement*. Perth: Government Printer.
Northwest Territories Department of Education. (1991). *Our students, our future: A planning framework*. Yellowknife, NT: Queen's Printer.
Ronkowski, S., & Iannaccone, L. (1989). *Socialization research in administration, graduate schools and other professions: The heuristic power of van Gennep and Becker models*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
Sergiovanni, T.J. (1984). Leadership and excellence in schooling. *Educational Leadership 41*(5), 4-13.
Spradley, J.P. (1979). *The ethnographic interview*. New York: Holt, Rinehart & Winston.
van Gennep, A. (1960). *The rites of passage*. Chicago: University of Chicago Press.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 337-353*

Genevieve Marie Johnson
University of Alberta

Undergraduate Student Attrition: A Comparison of the Characteristics of Students Who Withdraw and Students Who Persist

A recent Commission of Inquiry on Canadian University Education reported that from the crude data available "it would appear that 42% of full-time students who entered a university in 1985 failed to get a degree from that university within five years." One hundred sixty-three undergraduate students who were required to withdraw by their university, 109 undergraduates who had withdrawn by completing the necessary withdrawal forms, 226 undergraduates who withdrew from university by simply not returning as anticipated by the Registrar, and 153 students continuing undergraduate programs were interviewed regarding demographic, academic, financial, personal, and learning characteristics and experiences. A comparison of the characteristics of these groups is provided in this article.

Selon des données brutes récentes qui proviennent de la Commision d'Enquête sur le Status de l'Éducation dans les Universités Canadiennes "parait-il que 42% des étudiant(e)s inscrit(e)s à plein temps en études universitaires en 1985 auraient failli d'obtenir un diplôme de cette même université pendant les cinq ans suivant leur inscription initiale." On a passé à l'entrevue: 163 étudiant(e)s qui ont été oubligé(e)s de se retirer de leur université, 109 étudiant(e)s qui se sont retiré(e)s en complétant les formulaires de désistement nécessaires, 226 étudiant(e)s qui, tout simplement, ne sont pas revenu(e)s aux études tel que prévu par le secrétaire et l'archiviste de l'université, et, 153 étudiant(e)s qui poursuivaient toujours leurs études, afin de s'informer à propos de leurs expériences personelles d'apprenant(e)s ainsi que de leurs caractéristiques personelles, démographiques, académiques, et économiques. Dans cet article on fait la comparaison des caractéristiques de ces différents groupes.

The recent *Commission of Inquiry on Canadian University Education* (Smith, 1991) reports that from the crude data available "it would appear that 42% of full-time students who entered a university in 1985 failed to get a degree from that university within five years" (p. 105). In the Commission's opinion, however, "the statistics are less important than the general lack of interest around the subject" (p. 105).

Apparently, there is a pervasive Canadian attitude that university simply is not for everyone and that it is only natural that some students discover their lack of interest or lack of suitability after a year or two of undergraduate study. Indeed, university attendance may lead individuals to determine their likes and dislikes and discover occupations compatible with interests and abilities. This process of self-discovery, however, may prove costly for some individuals. In most cases individuals are, at least partially, removed from the work force while they attend university, resulting in financial loss. Equally, many

Genevieve Marie Johnson is a lecturer in the Department of Educational Psychology. A recent Social Sciences and Humanities Research Council of Canada Doctoral and Postdoctoral Fellow, her major research focus is concerned with learners at risk.

individuals leave their homes and communities to attend university, resulting in considerable disruption of their personal lives. One is forced to consider whether university attendance is the most efficient means by which individuals can determine their interests in and suitability for a university education.

Beyond concern for individual students, the consequences of high rates of student attrition are of concern to educational institutions. Canada has among the highest postsecondary education participation rates in the world, 27.5% of traditionally relevant age groups. Preceded only by the United States, 15% of Canadians hold a university degree (MacGregor, 1992). If Canada is to maintain, and perhaps even improve, this record of postsecondary participation and completion, careful consideration must be given to students who attend but fail to complete undergraduate programs of study. High rates of undergraduate student attrition may reflect inefficient utilization of continually dwindling resources. For example, the simple costs of manual and electronic record keeping place strain on administrative budgets and personnel. Instructors are plagued with balancing formal class lists with actual student attendance. An all too familiar instructor and/or administrator comment is "He or she must have dropped out."

Despite the extensive literature that addresses issues of student attrition, much remains unknown about the longitudinal process and complex interplay of forces that give rise to undergraduate student attrition. Students who withdraw are often portrayed as having a particular personality profile or as lacking important attributes prerequisite to university success (Blanchfield, 1971; Munro, 1981; Ott, 1988; Pascarella, Smart, & Ethington, 1986). As Gilbert (Smith, 1991) observes, however, "there is no real mystery about attrition; students re-enroll when they are having an exciting, substantive learning and personal growth experience that they can relate to their future development and success" (p. 106).

We need, then, specific information concerning Canadian undergraduate students who leave their universities prior to degree completion and, more generally, a model of the process of institution-based and student-based undergraduate withdrawal. Such a model, coupled with valid information, would be policy relevant, not merely of academic interest.

A Model of Undergraduate Student Attrition

A wide range of student characteristics have been empirically associated with undergraduate attrition. Academic factors such as limited hours of study, inefficient study skills, absenteeism, marginal academic prerequisite competences, and vague educational goals have all been linked to students prematurely leaving institutions of higher education (De Rome & Lewin, 1984; Getzlaf, Sedlacek, Kearney, & Blackwell, 1984; Grosset, 1991; Johnes, 1990; Van Overwalle, 1989). In addition, personal variables such as poor health, financial stress, employment, family responsibilities, gender, age, ethnicity, and lack of outside encouragement have been associated with college withdrawal (Boutsen & Colbry, 1991; Braxton, Brier, & Hossler, 1988; Ethington, 1990; Fox, 1986; Johnson & Miller, 1993; Johnson et al., 1992; Lang, Dunham, & Alpert, 1988; Mallette & Cabrera, 1991; Metzner & Bean, 1987; Moline, 1987; Nora, 1987; Theophilides, Terenzini, & Wendell, 1984).

In addition to specific student attributes, there is some evidence that campus integration is more characteristic of students who are successful in completion of their college programs than of students who are not (Allen & Nelson, 1989; Bers & Smith, 1991; Stage, 1989; Terenzini & Wright, 1987). Campus integration is defined as the extent to which students integrate into campus social life (e.g., make friends, join campus clubs and organizations). As well, institutional variables such as instructor behavior, number of students enrolled, and student support services have been associated with differential student postsecondary persistence (Bauer, 1981; Johnes & Taylor, 1989; Pascarella & Terenzini, 1983; Stampen & Cabrera, 1986).

With respect to undergraduate students who withdraw, a distinction must be made between those who voluntarily withdraw and those who are required to withdraw by their institution (Hayes, 1977; Simpson, Baker, & Mellinger, 1980; Tinto, 1987). In this regard, "the label dropout is one of the more frequently misused terms in our lexicon of educational descriptors" (Tinto, 1987, p. 3). It is used to describe the actions of all students who leave university without a degree, regardless of the reasons or circumstances that precipitated their departure. Graphically presented in Figure 1, undergraduate withdrawal is conceptualized as the consequence of either an institutional or a personal decision. Institutional decisions are generally based on inadequate student academic performance, although in a few cases institutions require students to withdraw for reasons of misconduct (e.g., plagiarism, cheating).

Personal withdrawal decisions are far more complex. Students make the decision to withdraw from undergraduate programs on the basis of two main factors, summarized in Figure 1 as Academic Performance and Psychological State. Students have a perception of the quality of their academic performance and evaluate that perception against a personal standard. Each student is characterized by a psychological state influenced by, among other things, campus integration and societal forces such as perceived employment options. Student academic performance and psychological state at the most fundamental level are the consequence of the relationship between student academic or personal characteristics and institutional factors. That is, each university student is characterized by a unique combination of academic (e.g., study skills, academic prerequisites) and personal (e.g., health, finances, family responsibilities) attributes. This unique combination of characteristics interacts with institutional variables such as course availability, instructor behavior, and student support services. It is the interaction between an individual and an institution that results in student academic performance and psychological state, directly or indirectly the bases for both institutional and personal withdrawal decisions.

Statement of the Problem

Undergraduate student attrition requires investigation both on behalf of students and on behalf of universities. How do students who withdraw from undergraduate programs differ from students who persist? What are the factors associated with undergraduate student withdrawal and persistence? How do students who voluntarily withdraw differ from students who are required to withdraw by the institution? What actions might universities take to increase undergraduate student retention?

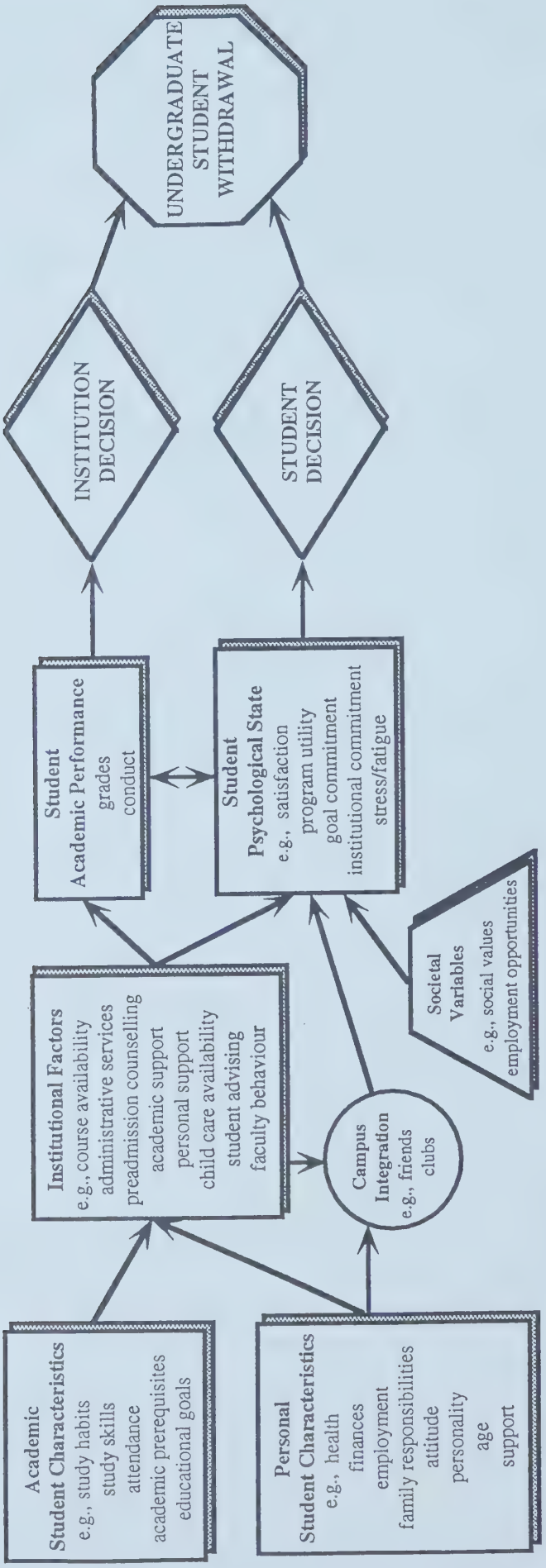


Figure 1. A model of undergraduate student attrition.

Methods and Procedures

Subjects

Information obtained from the Office of the Registrar of a large Western Canadian university showed that from mid-September 1991 to mid-September 1992, excluding students auditing all their courses, visiting students, special students, and unclassified students, approximately 10% of the undergraduate population left the institution without a degree. This one-year nondegree departure rate is similar to those recently reported by the Ontario University Registrars' Association (Gilbert & MacLean, 1992). From this approximate 10% of the undergraduate population, 674 names were randomly generated (the number of subjects selected was directly related to financial resources available). Of these 498 were contacted and agreed to participate (some of those on the list were not located due to disconnected telephones or no telephone number provided), resulting in a response rate of 74%. Once contacted students were asked to specify the nature of their university withdrawal. Three distinct types of university withdrawal emerged:

1. *Students who were required to withdraw by the institution*: 163 of the 498 withdrawn students (32.7% of the sample of students) did so at the request of the institution.
2. *Students who voluntarily withdrew and completed the appropriate withdrawal forms*: 109 of the 498 withdrawn students (21.9% of the sample of students) followed official university protocol by completing the necessary withdrawal forms.
3. *Students who did not continue/return as anticipated and who did not complete the necessary withdrawal forms*: 226 of the 498 withdrawn students (34.7% of the sample of students) did not follow official withdrawal protocol, that is, they simply did not return/continue as anticipated by the Registrar.

For purposes of comparison a corresponding sample of students continuing undergraduate programs was obtained. A random list of 167 continuing undergraduate students (number of subjects selected directly related to financial resources available) was generated (i.e., individuals in attendance at the institution in September 1991 who did not graduate and who were continuing undergraduate programs in September 1992). Of these 167 continuing undergraduate students, 153 were contacted and agreed to participate (a few individuals from the initial random list were not located), resulting in a response rate of 92%.

Withdrawn and continuing student samples were distributed across 18 faculties and programs. Large faculties were most frequently represented and, correspondingly, small faculties/programs were marginally represented. Summarized in Table 1, for example, Faculty of Arts students constituted 20.7%, Faculty of Education students 17.5%, and Faculty of Science students 33.9% of the withdrawn sample. With respect to the continuing student sample, 0.7% were registered in the Faculty of Law, 0.7% were registered in Dental Hygiene, and 0.7% were registered in Medical Laboratory Sciences. (Small individual faculty sample size renders faculty-specific generalization of findings questionable.)

Student withdrawal was progressively less likely as year of program increased. As presented in Table 2, 40.4% of the withdrawn students were in their

first year of undergraduate study, 30.3% in their second year, 17.7% in their third year, and 11.6% in their fourth year of study. Because of initial sampling procedures, first year students were marginally represented in the continuing student group. Recall that for comparison purposes, continuing students were drawn from a sample that was in attendance in September 1991, that did not graduate, and that was in attendance in September 1992. The nine first-year students in the continuing sample were reportedly in their first-year of undergraduate studies for two consecutive years; they were part-time students. Males and females were equally represented in the withdrawn student sample (49.4% and 50.6% respectively) and equally represented in the continuing student sample (49.7% and 50.3% respectively).

Data Collection

A telephone interview questionnaire was developed that asked students a number of questions concerned with demographic, academic, financial, personal, and learning characteristics and experiences. Questionnaire items were forced-choice, that is, questions were answered by selecting from a set of choices (e.g., yes-no-maybe; never-sometimes-often).

Ten telephone interviewers were hired and trained. Students were contacted during February and March 1993. Interviews were conducted on all days of the week, with contact being most frequent on Saturday and Sunday. Interviewers often had to make repeated attempts to contact students. For the continuing student group, average number of callbacks was 2.6 (range 0-12) and for the withdrawn student group 3.2 (range 0-18).

Table 1
Faculty Distribution of Continuing and Withdrawn Samples

Faculty	*Continuing Students*		*Withdrawn Students*	
	Frequency (n = 153)	*Percentage*	*Frequency (n = 498)*	*Percentage*
Agriculture and Forestry	2	1.3	16	3.2
Arts	27	17.6	103	20.7
Business	14	9.2	26	5.2
Dentistry and Dental Hygiene	3	2.0	1	0.2
Education	26	17.0	87	17.5
Engineering	17	11.1	49	9.8
Faculté St. Jean	2	1.3	14	2.8
Home Economics	2	1.3	5	1.0
Law	1	0.7	2	0.4
Medicine and Medical Laboratory Sciences	6	4.0	4	0.8
Native Studies	2	1.3	1	0.2
Nursing	2	1.3	13	2.6
Pharmacy	6	3.9	1	0.2
Physical Education and Recreation	4	2.6	6	1.2
Rehabilitation Medicine	6	3.9	1	0.2
Science	33	21.6	169	33.9

Table 2
Year of Program Distribution of Continuing and Withdrawn Samples

Year of Program	Continuing Students		Withdrawn Students	
	Frequency (n = 153)	Percentage	Frequency (n = 498)	Percentage
First Year	9	5.9	201	40.2
Second Year	46	30.1	151	30.3
Third Year	51	33.3	88	17.7
Fourth Year	45	29.4	58	11.6
Fifth Year	2	1.3	0	0.0

Data Analysis

The four student groups (163 students who were required to withdraw by the institution, 109 students who withdrew by completing the necessary withdrawal forms, 226 students who withdrew by simply not returning as anticipated by the Registrar, 153 students continuing their undergraduate programs of study) were compared in terms of proportional (e.g., percent of sample that borrowed money from the Student Finance Board) or mean (e.g., average amount borrowed from the Student Finance Board) questionnaire responses. Analyses of Variance were computed in order to determine the significance of differences between groups. For purposes of reporting and discussing results, questionnaire items were organized into meaningful categories (e.g., background/entering student characteristics, student academic behavior and performance, student finance and employment, campus integration, personal characteristics, and evaluation of learning experiences).

Results and Discussion

Table 3 presents four-group comparisons for student background characteristics, that is, characteristics that students possessed on university entry. Students who were required to withdraw were on average younger than the continuing comparison group, 21.4 years and 23.1 years respectively. Students who officially and unofficially voluntarily withdrew from university, on the other hand, were older than the continuing comparison group (F=28.737, p<.001, df=2). Students who withdrew from university, either officially or by simply not returning, were less likely to have attended university directly after high school than were students required to withdraw by the institution (F=15.358, p<.001, df=2). Youth and inexperience characterized students who were required to withdraw by the institution. Relatively greater maturity characterized students who voluntarily withdrew or simply did not return to university.

Significant differences across the four groups were found in student-reported grade 12 marks (F=5.117, p<.01, df=2). On average, students who were required to withdraw reported grade 12 averages of 78.7%. Students who voluntarily withdrew by completing the necessary withdrawal forms as well as those who did not return reported grade 12 averages of 81.1%. Students continuing their undergraduate programs reported grade 12 averages of 83.1%. Students who officially or unofficially voluntarily withdrew were five to six

times more likely than students who were required to withdraw to report entering university as nonmatriculated students (F=5.151, p<.01, df=2). Entering university as a nonmatriculated student was most frequently associated with student-initiated withdrawal decisions. Entering university as a matriculated student, on the other hand, was most frequently associated with institution-initiated withdrawal decisions. Apparently, there is a set of attributes that typically characterizes undergraduate students most likely to persist with their university programs. This potentially persistent undergraduate will not be too young nor too old. He or she will have been employed prior to university attendance, but will not have worked for too long. This hypothetically persistent university student has a relatively high grade 12 average. As Canadian universities move with the current tide of lifelong learning, consideration must be given to the instructional accommodation of students who deviate from what has traditionally been understood as the typical undergraduate student.

The four student groups differed from one another in terms of academic behavior and performance (Table 4). Approximately half the students who were required to withdraw by the institution reported frequently missing classes, while fewer than one third of students who officially or unofficially voluntarily withdrew reported the same (F=15.791, p<.001, df=2). Correspondingly, 22% of those students who were required to withdraw reported frequently failing to submit course assignments, while 7.3% of the group that voluntarily withdrew by completing the necessary withdrawal forms and 4.8%

Table 3
Across-group Comparison of Background/Entering Student Characteristics

Abbreviated Item	*Student Group*			
	Required to withdraw (n = 163)	*Voluntarily[a] withdrew (n = 109)*	*Did not[b] return (n = 226)*	*Continuing (n = 153)*
Age in years	21.4	26.6	25.7	23.1
	(3.25)	(8.38)	(7.24)	(5.73)
Attended university directly after high school	63.2%	32.1%	42.0%	56.9%
Grade 12 average	78.7%	81.1%	81.1%	83.1%
Employed full-time prior to university attendance	20.2%	46.8%	40.3%	24.8%
Number of years employed	4.1	9.3	8.5	6.0
	(4.71)	(8.43)	(7.34)	(5.50)
Entered as non-matriculation student	3.1%	16.5%	16.8%	9.2%
Father has university degree	29.4%	34.9%	32.3%	37.3%
Mother has university degree	23.9%	26.6%	22.1%	30.1%
Enrolled without clear career goals	41.1%	32.1%	39.8%	42.5%

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

of those students who simply did not return reported the same. With respect to the continuing comparison group, only 2.6% reported frequently not submitting course assignments (F=24.817, p<.001, df=2). Falling behind in course related readings was equally characteristic of students who were required to withdraw and of students who were continuing undergraduate programs of study but significantly different from the groups of students who officially and unofficially voluntarily withdrew from university (F=15.587, p<.001, df=2). Students who were required to withdraw showed a clear pattern relative to the other student groups of depressed academic performance. Student-reported grade point averages showed significant differences across the four student groups (F=142.754, p<.001, df=2). In addition, 97% of the sample of students who were required to withdraw by the university reported having failed on average 2.7 courses during their undergraduate program of studies.

In general, students who voluntarily withdrew from university by simply not returning were registered in fewer courses than any of the other student groups. Both students who were required to withdraw and those who voluntarily withdrew were on average registered in fewer courses than students in the continuing comparison group (F=13.993, p<.001, df=2). Although it might be assumed that students with lighter course loads would have more time to

Table 4
Across-group Comparison of Student Academic Behavior and Performance

Abbreviated Item	*Student Group*			
	Required to withdraw (n = 163)	*Voluntarily[a] withdrew (n = 109)*	*Did not[b] return (n = 226)*	*Continuing (n = 153)*
Often missed classes	52.1%	31.2%	30.1%	17.6%
Often did not submit assignments	22.0%	7.3%	4.8%	2.6%
Often fell behind in readings	69.3%	35.7%	41.2%	63.1%
Grade point average	4.5	6.1	6.4	7.0
	(1.11)	(1.27)	(1.09)	(0.85)
Have poor study skills	68.1%	25.7%	29.2%	26.8%
Have inefficient time-management	53.4%	33.0%	23.9%	26.8%
Average number of courses	4.6	4.2	3.7	5.1
registered in	(1.43)	(1.62)	(1.66)	(2.32)
Ever withdrawn from a course	74.8%	83.5%	68.6%	58.8%
Average number of course withdrawals	2.2	2.9	2.6	2.5
	(1.43)	(2.00)	(2.22)	(1.88)
Ever placed on academic probation	39.3%	19.3%	21.2%	14.4%
Ever failed university course	97.0%	36.7%	37.6%	23.5%
Average number of courses failed	2.7	1.7	2.0	1.9
	(1.42)	(0.97)	(1.74)	(1.11)

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

Table 5
Across-group Comparison of Student Employment and Finance Characteristics

Abbreviated Item	*Student Group*			
	Required to withdraw (n = 163)	*Voluntarily[a] withdrew (n = 109)*	*Did not[b] return (n = 226)*	*Continuing (n = 153)*
Employed last three months of university attendance	47.9%	48.6%	53.5%	39.2%
Hours per week employed	19.9	20.0	23.2	16.2
	(12.49)	(10.69)	(13.83)	(10.24)
Borrowed from Student Finance Board	42.3%	55.0%	51.8%	58.2%
Average amount borrowed	$6583	$8865	$8513	$10,538
	(7291)	(7564)	(6127)	(6756)
Enough money to meet basic needs	89.6%	89.9%	88.5%	83.7%
Experience financial problems	18.4%	22.0%	23.9%	28.1%

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

concentrate on course requirements than students with heavy course loads, such did not appear to be the case. Instead, it appeared that students who were taking a full course load were more likely to persist with their undergraduate studies than students who were, for example, registered as part-time students. Indeed, students registered in less than a full course load may be burdened with employment and familial demands. Interestingly, students who voluntarily withdrew from university by completing the required withdrawal forms were more likely than any of the other student groups to report having withdrawn from a course or courses; 83.5% stated that they had withdrawn from an average of 2.9 courses (F=4.338, p<.05, df=2). Because they complied with university procedures, it may be that students who voluntarily withdrew by completing withdrawal forms were more accommodating to university protocol than were students who simply did not return to university as anticipated by the Registrar. Equally, students who officially withdrew from university by completing the necessary forms were in control of their actions, just as students who withdrew from individual university courses demonstrated control of the actions.

Table 5 presents across-group comparisons for questionnaire item responses concerned with student employment and general financial situation. During university attendance, students who withdrew were not significantly more often employed than were students who persisted. Relative to students who withdrew from university, students who continued their academic programs were not significantly more likely to report having borrowed money from the Student Finance Board and were not significantly more indebted.

However, students who persisted with their undergraduate programs generally reported greater financial hardship than did students who withdrew. As consistently reported in the literature, financial situation does not appear related to student attrition from university. Nevertheless, it is possible that individuals vary in their capacity to tolerate financial hardship. Or perhaps tolerance of financial hardship is associated with other variables such as student age, level of independence, and extent of personal commitment to program continuation. Regardless, indebtedness and unemployment were not significantly more characteristic of students who continued than of students who withdrew from undergraduate programs of study.

Campus integration questionnaire item responses are summarized and compared in Table 6. Students who continued undergraduate programs did not necessarily show a clear pattern, relative to students who withdrew from university, of more complete integration into campus life. Relatively large proportions of both continuing and withdrawn undergraduate students reported incidence of campus isolation. For example, approximately half of those students who continued their undergraduate programs reported being actively involved in campus clubs and organizations; approximately one third of students who withdrew reported the same. Students who withdrew were not significantly more likely to report feeling lonely at university than were students who persisted in their undergraduate studies. Correspondingly, students who withdrew were not significantly more likely than students in the continuing comparison group to report feeling alienated while attending university. Equally, student-reported experience or perception of on-campus discrimination was not significantly different across the four student groups. The overwhelming majority of students in the continuing group, however, reported that leaving university prior to degree completion would be an unpleasant experience (91.5%). Fewer than half of those students required to withdraw by the institution reported that their university departure was an unpleasant experience (47.9%). Fewer than one third of those students who withdrew by completing the necessary withdrawal forms viewed the withdrawal experience as unpleasant (31.1%). Fewer than one quarter of the students who withdrew from university by not returning recounted their withdrawal experience as unpleasant (22.5%). Such differences did not occur by chance (F=22.802, p<.001, df=2). Apparently, students who have control of their university careers (i.e., officially or unofficially voluntarily withdraw) have a more positive interpretation of the university withdrawal experience than do students under the control of the institution (i.e., required to withdraw by the university).

As summarized in Table 7, the four student groups differed from one another on a number of personal attributes. The group of students who voluntarily withdrew by completing the appropriate forms showed the lowest rate relative to the other three groups of academic confidence; 7.3% reported that they felt unable to succeed at university (F=3.227, p<.05, df=2). Students who withdrew were not significantly more likely as a group and in relation to students in the continuing comparison group to be characterized by family problems, but were significantly more likely to report personal problems (F=5.091, p<.05, df=2). Students who were required to withdraw by the institu-

Table 6
Across-group Comparison of Student Integration into Campus Life

Abbreviated Item	*Student Group*			
	Required to withdraw (n = 163)	*Voluntarily[a] withdrew (n = 109)*	*Did not[b] return (n = 226)*	*Continuing (n = 153)*
Belonged to university clubs/organizations	36.8%	28.4%	36.3%	51.0%
Average hours per week at club activities	5.2 (12.49)	4.5 (3.72)	6.8 (8.13)	6.2 (7.09)
Felt lonely at university	19.0%	15.6%	20.8%	7.2%
Felt alienated at university	25.3%	18.3%	18.1%	5.2%
Felt university was too impersonal	63.1%	44.0%	47.4%	33.3%
Had close university student-friends	69.9%	70.7%	60.2%	88.2%
Felt homesick	13.5%	7.4%	8.8%	11.8%
Felt discriminated against	3.7%	7.4%	9.3%	1.3%
Leaving university was/would be unpleasant	47.9%	31.1%	22.5%	91.5%

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

tion were four times more likely than students continuing undergraduate programs to define their social lives as excessive (F=24.208, p<.001, df=2). Not one student in the continuing comparison group reported attending university as a response to outside pressure. In contrast, approximately 5% of those students who personally initiated their undergraduate withdrawal reported experiencing outside pressure to attend university; 9.2% of those students who withdrew at the request of the institution reported the same (F=5.170, p<.01, df=2). Lack of familial support was also more characteristic of students who withdrew from university than of students who persisted.

Relative to the continuing group, marriage was significantly more characteristic of students who initiated their own withdrawal from university and less characteristic of students whose university withdrawal was initiated by the institution (F=18.283, p<.001, df= 2). Approximately 10% of the continuing comparison group reported having children; students who were required to withdraw were half as likely to be parents; students who voluntarily withdrew as well as those who simply did not return were more than twice as likely to be parents (F=10.723, p<.001, df=2). Apparently, marriage and parenthood are risk factors for undergraduate student withdrawal from university. Being married with children increases the probability of student-initiated undergraduate withdrawal; being single without children increases the probability of institution-initiated undergraduate withdrawal.

The four student groups showed differences with respect to evaluation of their learning experiences at university (Table 8). Among students who

Table 7
Across-group Comparison of Student Personal Characteristics

Abbreviated Item	Student Group			
	Required to withdraw (n = 163)	Voluntarily[a] withdrew (n = 109)	Did not[b] return (n = 226)	Continuing (n = 153)
Felt unable to succeed at university	5.5%	7.3%	2.7%	3.9%
Had family problems	21.5%	18.2%	15.9%	7.8%
Had personal problems	33.7%	30.2%	23.0%	9.1%
Socialized excessively	26.3%	13.8%	8.0%	7.9%
Outside pressure to attend university	9.2%	4.6%	4.9%	0.0%
Family not support decision to attend	4.9%	2.7%	3.1%	1.3%
Never married	93.3%	69.7%	65.0%	85.6%
Have children	4.9%	21.1%	20.4%	9.8%
Average number of children	1.3	2.6	2.0	1.9
	(0.44)	(1.09)	(1.18)	(1.37)
Had child care problems	1.8%	4.6%	3.5%	2.6%

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

withdrew from university, more notably for students who were required to withdraw than for those who voluntarily withdrew, there existed a perception, more apparent than in the continuing comparison group, that instructors were unavailable (F=16.800, p<.001, df=2) and that the quality of instruction was poor (F=4.941, p<.01, df=2). Students who were required to withdraw were three to four times more likely to report not enjoying their undergraduate classes than were students continuing undergraduate programs of study (F=7.215, p<.01, df=2).

Approximately half the students who withdrew felt that their undergraduate programs were not developing skills necessary for employment. Although 11.8% of the continuing comparison group reported the same perception, such group differences may have occurred by chance. A similar, although less pronounced, pattern was evident with respect to students' evaluation of their learning of specific competences. For example, 30.1% of required to withdraw students reported that they had not learned oral expression while attending university, and 21.5% reported that they had not learned written expression. In contrast, 11.1% of the continuing comparison group reported that they had not learned oral expression while attending university and 11.7% reported that they had not learned written expression. Although these group differences are of interest, across-group significance was not established.

Theoretical Contributions and Implications for Practice

Both validation and reconsideration of the proposed Model of Undergraduate Student Attrition (Figure 1) are reflected in the current findings. Campus

Table 8
Across-group Comparison of Student Evaluation of Learning Experiences

Abbreviated Item	*Student Group*			
	Required to withdraw (n = 163)	*Voluntarily[a] withdrew (n = 109)*	*Did not[b] return (n = 226)*	*Continuing (n = 153)*
Felt instructors were unavailable	38.0%	20.1%	22.1%	11.8%
Felt poor quality of instruction	29.5%	21.1%	25.7%	14.4%
Unable to take desired courses	19.6%	22.1%	27.4%	24.2%
Did not enjoy classes	19.6%	19.3%	11.5%	5.9%
University attendance is/was a bad experience	11.8%	8.3%	7.1%	2.7%
Did not learn verbal expression	30.1%	21.1%	28.8%	11.1%
Did not learn written expression	21.5%	18.4%	19.0%	11.7%
Did not learn employment skills	50.3%	44.9%	45.5%	11.8%
Program was wrong decision	20.8%	20.2%	17.6%	4.0%

Note. As applicable, figures reflect group mean (and standard deviation) or proportion of the group responding in the affirmative to the questionnaire item.
[a]student-initiated official university withdrawal (completed voluntary withdrawal forms)
[b]student-initiated unofficial university withdrawal (did not complete voluntary withdrawal forms but, rather, simply did not return to/continue university as anticipated by the Registrar).

integration, at least as measured by the questionnaire items (e.g., belonged to university clubs/organizations, felt lonely at university, felt alienated at university, felt discriminated against), did not appear specifically related to undergraduate student withdrawal or persistence. Institutional commitment, measured by an affirmative response to the questionnaire item leaving university was/would be unpleasant, was clearly more characteristic of students who persisted with undergraduate programs than of students who withdrew. Certain academic student characteristics proposed in Figure 1 were found to relate to student attrition whereas others were not. For example, varying levels of academic prerequisite competences were differentially associated with undergraduate student attrition. Correspondingly, certain personal student characteristics proposed in Figure 1 were found to relate to student attrition whereas others were not. For example, student age was significantly related to student withdrawal, but student finance did not appear to be associated with withdrawal. Student grades were obviously related to institution-initiated university withdrawal decisions; student psychological state was most strongly associated with student-initiated university withdrawal.

Students who were required to withdraw by the institution were different from students who voluntarily withdrew. Both groups of students, however, differed from the typical continuing undergraduate student. A composite profile of the continuing undergraduate student emerged from the data. Deviation from this profile was associated with increased probability of university withdrawal. Relative to the typical continuing university student, students who were required to withdraw by the institution were younger. Students who voluntarily withdrew, on the other hand, were older than the typical continu-

ing undergraduate student. Students who withdrew on average entered university with slightly lower grade 12 averages than did students who persisted in undergraduate programs. An increased focus on preadmission counseling services may serve to reduce undergraduate student attrition. Prior to admission, students should be advised of the characteristics most typically associated with student persistence in undergraduate programs. In some cases, students may benefit from delayed admission to university in favor of relevant work experience. Preadmission awareness of the characteristics more common of students who withdraw than of students who persist may result in student as well as institution proactive action. For example, increased focus on helping students deal with personal problems may serve to increase retention.

In general, students who voluntarily withdrew did not report having less efficient time management and study skills than did continuing undergraduate students. Students who were required to withdraw, however, on average reported less efficient time management and study skills than did students continuing undergraduate programs. Relative to continuing students, students who voluntarily withdrew reported somewhat lower academic performance, but students who were required to withdraw reported considerably lower academic performance. Postadmission academic support services may serve to increase undergraduate student program completion. Some students may benefit from remedial reading and writing and perhaps tutorial support for particularly problematic or challenging courses.

Students who were required to withdraw as well as those who voluntarily withdrew did not significantly differ from students who were continuing undergraduate programs in terms of patterns of employment and general financial situations. As generally reported in the literature, financial situation is not related to student attrition from university. Nevertheless, a relatively large proportion of the continuing undergraduate students surveyed reported that they were experiencing financial hardship in order to pursue university careers. Although not predictive of undergraduate attrition, such information should not be ignored by university administrators and student services personnel.

Students who withdrew, particularly those who were required to withdraw, were more likely than students persisting with undergraduate programs to report the perception that instructors were unavailable and that the instruction they received was poor. An increased commitment to excellence in undergraduate teaching with specific focus on students who are academically weak may serve to increase student retention. Creative means must be found to ensure that all undergraduate students obtain the individual instructional attention they require. For example, electronic teacher-student interaction (Johnson & Johnson, in press) appears to hold considerable promise for individualizing instruction at the undergraduate level.

Acknowledgments

Special thanks to Dr. Harvey Krahn and the University of Alberta Department of Sociology Population Research Laboratory for conducting the telephone questionnaire survey, University of Alberta Registrar Brian Silzer for assistance with subjects and sampling, University of Alberta Student Services Unit Heads for input regarding questionnaire development, and Mr. Roy Jaffray

(University of Alberta Centre for Research in Applied Measurement and Evaluation) for assistance with data manipulation and analysis.

References

Allen, D.F., & Nelson, J.M. (1989). Tinto's model of college withdrawal applied to women in two institutions. *Journal of Research and Development in Education, 22*, 1-11.

Bauer, W.K. (1981). Strategies for recruiting and retaining the nontraditional student. *College Student Journal, 15*, 234-238.

Bers, T.H., & Smith, K.E. (1991). Persistence of community college students: The influence of student intent and academic and social integration. *Research in Higher Education, 32*, 539-556.

Blanchfield, W.C. (1971). College dropout identification: A case study. *Journal of Experimental Education, 40*, 1-4.

Boutsen, F.R., & Colbry, S.L. (1991). Single-parent college students' attributions toward academic success. *Psychology in the Schools, 28*, 360-364.

Braxton, J.M., Brier, E.M., & Hossler, D. (1988). The influence of student problems on student withdrawal decisions: An autopsy on the "autopsy" studies. *Research in Higher Education, 28*, 241-253.

De Rome, E., & Lewin, T. (1984). Predicting persistence at university from information obtained at intake. *Higher Education, 13*, 49-66.

Ethington, C.A. (1990). A psychological model of student persistence. *Research in Higher Education, 31*, 279-293.

Fox, R.N. (1986). Application of a conceptual model of college withdrawal to disadvantaged students. *American Educational Research Journal, 23*, 415-423.

Getzlaf, S.B., Sedlacek, G.M., Kearney, K.A., & Blackwell, J.M. (1984). Two types of voluntary undergraduate attrition: Application of Tinto's model. *Research in Higher Education, 20*, 257-268.

Gilbert, S., & MacLean, J. (1992, June). *Measuring attrition/retention rates in Ontario universities.* Paper presented at the Annual Conference of the Association of Registrars of Universities and Colleges of Canada, Moncton.

Grosset, J.M. (1991). Patterns of integration, commitment, and student characteristics and retention among younger and older students. *Research in Higher Education, 32*, 159-178.

Hayes, S.C. (1977). Dropouts' dissatisfaction with university. *Australian Journal of Education, 21*, 141-149.

Johnes, J. (1990). Determinants of student wastage in higher education. *Studies in Higher Education, 15*, 87-99.

Johnes, J., & Taylor, J. (1989). Undergraduate non-completion rates: Differences between UK universities. *Higher Education, 18*, 209-225.

Johnson, G.M., et al. (1992). *Student child care.* Council on Student Life Report No. 9. University of Alberta, Edmonton, Canada.

Johnson, G.M., & Miller, P.J. (1993). The characteristics of university student-parents: The impact of parenthood on the university experience. In *Postsecondary Education in Transition: Transactions of the Annual Conference of the Canadian Society for the Study of Higher Education* (pp. 202-206). Office of Institutional Research and Planning, University of Ottawa.

Johnson, J.A., & Johnson, G.M. (in press). Automating the professional journal technique for computer science education: A case study. *Journal of Research on Computing in Education.*

Lang, G., Dunham, R.G., & Alpert, G.P. (1988). Factors related to the academic success and failure of college football players: A case of mental dropout. *Youth and Society, 20*, 209-221.

MacGregor, K. (1992, September 25). Drop-out rates damn growth. *Times Higher Education Supplement*, 1, 10-11.

Mallette, B.I., & Cabrera, A.F. (1991). Determinants of withdrawal behavior: An exploratory study. *Research in Higher Education, 32*, 179-194.

Metzner, B.S., & Bean, J.P. (1987). The estimation of a conceptual model of nontraditional undergraduate student attrition. *Research in Higher Education, 27*, 15-38.

Moline, A.E. (1987). Financial aid and student persistence: An application of causal modeling. *Research in Higher Education, 26*, 130-147.

Munro, B.H. (1981). Dropouts from higher education: Path analysis of a national sample. *American Educational Research Journal, 18*, 133-141.

Nora, A. (1987). Determinants of retention among Chicano students: A structural model. *Research in Higher Education, 26*, 31-59.

Ott, M.D. (1988). An analysis of the predictors of early academic dismissal. *Research in Higher Education, 28*, 34-48.

Pascarella, E.T., Smart, J., & Ethington, C. (1986). Long-term persistence of two-year college students. *Research in Higher Education, 24*, 47-71.

Pascarella, E.T., & Terenzini, P.T. (1983). Predicting voluntary freshmen year persistence/ withdrawal behavior in a residential university: A path analytic validation of Tinto's model. *Journal of Educational Psychology, 75*, 215-226.

Simpson, C., Baker, K., & Mellinger, G. (1980). Conventional failures and unconventional dropouts: Comparing different types of university withdrawals. *Sociology of Education, 53*, 203-214.

Smith, S.L. (1991). *Commission of inquiry on Canadian university education*. Ottawa: Association of Universities and Colleges of Canada.

Stage, K.F. (1989). Reciprocal effects between academic and social integration of college. *Research in Higher Education, 30*, 517-530.

Stampen, J.O., & Cabrera, A.F. (1986). Exploring the effects of student aid on attrition. *Journal of Financial Aid, 16*, 28-39.

Terenzini, P.T., & Wright, T.M. (1987). Influences on students' academic growth during four years of college. *Research in Higher Education, 26*, 161-179.

Theophilides, C., Terenzini, P.T., & Wendell, L. (1984). Relation between freshmen-year experience and perceived importance of four-major educational goals. *Research in Higher Education, 20*, 235-252.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago, IL; University of Chicago Press.

Van Overwalle, F. (1989). Success and failure of freshman at university: A search for determinants. *Higher Education, 18*, 287-308.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 355-373*

B.E.J. Housego
University of British Columbia

How Prepared Were You to Teach? Beginning Teachers Assess Their Preparedness

Teachers are being challenged to alter substantially their professional practice to meet changes in schools and communities. Teacher educators are likewise called to be accountable for their practice. This follow-up study used Q-sort methodology to examine perceptions of first-year teachers' preparedness to teach and explored the effects of both strengths and weaknesses in preparedness. Revisiting their teacher education program in light of a year's classroom experience, the teachers acknowledged some program strengths and offered some suggestions for changes and improvements.

De plus en plus, les enseignant(e)s sont défié(e)s à faire des modifications considérables à leur pratique professionnelle afin de faire face aux changements scolaires et communautaires contemporains. C'est également le cas pour ceux et celles qui oeuvrent à la formation des enseignant(e)s. En utilisant une méthodologie "Q-sort" on a pu examiner les perceptions vis-à-vis l'état de préparation envers l'enseignement des enseignant(e)s après leurs première année d'enseignement. On a aussi exploré les effets des forces et des faiblesses de cet état de préparation envers l'enseignement. En faisant un retour sur leur programme de formation professionnelle après un an en salle de classe, les enseignant(e)s ont pu reconnaître les points forts du programme de formation et ils/elles ont pu offrir quelques suggestions afin de l'améliorer et d'y apporter des changements.

Genesis of this Study

Eleven years ago, about the time when teacher educators were taking heart in the emergence of a more substantial knowledge base for teaching (Berliner, 1984), a committee began to redesign University of British Columbia teacher education programs. Four and a half years later, the new elementary and secondary programs, designed somewhat in the research-based or social efficiency tradition, were first offered. They were directed less to the technical version, which emphasizes skill training to a minimal mastery level, and more to the deliberative version, in which the goal is to develop the capability to deliberate about the "use of research-based skills with other factors, within a conception of teacher as decision-maker" (Zeichner, 1993, p. 5). The distinction between the technical and deliberative versions of research-based teacher education was introduced by Feinman-Nemser (1990).

The occasion of major program alteration presented a unique opportunity to study the changing process of educating teachers. Program changes included movement, almost entirely to a post-baccalaureate or consecutive preparation model and the adoption of a developmental practicum sequence starting with a series of regular half days, followed by a two-week orientation

Billie Housego is a member of the Department of Educational Psychology and Special Education in the Faculty of Education. She is also associated with the Centre for the Study of Teacher Education.

practicum and ending with a 13-week, single-placement, extended practicum. Several courses were designed in large, single-lecture format supplemented with small-group laboratories to permit follow-up discussion and coached skill development. Chief among these was a generic-skills teaching course called Principles of Teaching. A course to prepare faculty members and sponsor teachers to advise student teachers was also implemented.

Feelings of preparedness to teach was chosen as a research focus in expectation that one of the initial outcomes of improving the program would be an increasing recognition and endorsement of preparedness by student teachers (Bandura, 1986). An interdepartmentally generated repertoire of skills considered essential to a beginning teacher, one of the foundations for program change, provided the framework for the subsequently developed PREP Scale (Housego, 1990b). This scale was used to compare the feelings of preparedness to teach of students in the old and new programs (Housego, 1990a) and to monitor the development of feelings of preparedness to teach from term to term across programs (Housego, 1992a, 1992b).

In this study a selection of graduates of one program was followed into first year teaching. The PREP Scale was used to determine not how prepared to teach these beginning teachers felt prior to entering the profession, but rather how prepared to teach they actually had found themselves to be once in it. In subsequent interviews, first-year teaching and preparation for it were discussed, tempering an otherwise technical perspective with some personal, even perhaps moral, soundings (Goodlad, 1991).

Procedure

Participants

The participants in this study were 16 graduates of the 1989-1991 University of British Columbia Faculty of Education Two-Year Elementary Program, 12 females and four males ranging in age from 24 to 40 years. Although most had come directly into teaching after completing a prior degree, at least five had worked in other capacities or spent time at home parenting. These persons were not chosen from among a larger group as a sample; the study uses replication logic (Yin, 1989). They consented to participate in response to a letter of invitation, also sent to others easily accessible for interview but from whom no response was received. Their teaching assignments spanned the elementary grades, kindergarten to grade 7, in six lower mainland school districts. One taught educationally handicapped children; another's assignment was largely as a music specialist. Four were working as long-term or daily substitute teachers near the end of the 1991-1992 school year.

The Scale and the Method

The Student Teachers' Feelings-of-Preparedness to Teach Scale (PREP Scale) was originally designed to measure the degree to which student teachers felt prepared to perform a set of tasks central to teaching and applicable across grade levels and subject matter fields. Each item is stated so as to complete the sentence "I feel prepared to ..." and accompanied by a 7-point scale, from "almost completely unprepared" to "almost completely prepared." In repeated administrations it has been found to be highly reliable (between 0.95 and 0.97 based on Hoyt's coefficient, an index of item homogeneity) and valid for the

purposes of the studies in which it has been used (Housego, 1990a, 1990b, 1992a, 1992b).

The participants in the present study were asked to Q-sort the 50 PREP Scale items; that is, to engage in a comparative ranking process in which every task is compared with every other, sorting them into 11 categories of predesignated size according to the degree to which they found they *actually were* prepared to perform them during their first year of teaching. At the extremities of the Q-distribution the teachers placed the two tasks for which they found they were actually most and least prepared, then the three, then the four, then the five, and finally the six tasks for which they felt they were actually next most or next least prepared, leaving 10 tasks in the central category. It was then possible to isolate among the 50, the two, five, nine, or even 14 or 20 tasks for which individual teachers and subgroups found themselves most or least prepared.

Feelings of preparedness to perform these tasks had been monitored across the two-year teacher education program (Housego, 1992a). It was possible, therefore, to trace a history of developing feelings of preparedness to teach for the tasks for which each teacher now indicated he or she was actually most and least prepared and to discuss these patterns and their effects in subsequently scheduled interviews during the last two months of the 1991-1992 school term, either on school sites or on campus. The interview questions, which also explored the experience of first-year teaching and retrospectively examined the contribution of the teacher education program to preparedness for teaching, were refined by the members of a doctoral seminar in the Center for the Study of Teacher Education. Most members had recent teaching experience in one or another Canadian province or, in some instances, in other countries, and many were currently supervising student teachers.

Interview data were transcribed, producing documentation of the 16 answers to each interview question. Each set of answers was perused for commonalities that are summarized later in this article.

Report and Discussion of Findings

Q-Group Findings

Three distinct Q-groups emerged. They were defined according to the teaching tasks for which members found themselves most and least prepared, that is, tasks for which z-scores or standard deviation scores were above +1 and below –1 respectively. All 16 participants experienced the same two-year teacher education program including the commonalities of some large-group instruction and, of course, the variation in multisectioned courses and individual practicum placements. Substitute teachers or those who began employment as substitute teachers were represented in all three Q-groups.

Group 1. This largest subgroup consisted of eight women who as a group identified themselves on the basis of strength and relative weakness of preparedness for the tasks displayed in Table 1. The group found itself well prepared to relate to students and to engage in the more creative aspects of planning instruction; for example, in tasks such as integrating learning, enriching instruction, designing activities based on a collection of materials, and drawing subject matter from personal background.

As a group, these beginning teachers found themselves least prepared for assessment tasks. They also declared a lack of preparedness to evaluate cur-

riculum materials from the standpoint of current guidelines or in the instance of their being controversial as might be the case, for example, in programs of family living and sex education or units of work on politically sensitive economic issues. In addition, they found classroom management problematic, both the routine type such as monitoring the class, implementing routines, and making smooth transitions between lessons; and the complex type, such as understanding the causes of student misbehavior. This group, which identified its strength of preparedness in relationship skills and creative endeavors and not in management or assessment, may be seen to align itself somewhat with a nurturing approach to teaching. Elbaz (1992) explains how the central concerns of mothers as described by Ruddick (1989), preservation, growth, and the shaping of an acceptable child, are also concerns for teachers and require the invention of language and conceptual categories to permit us to talk about teachers' work and thinking in these terms without accusations of sentimentality of the sort experienced by Nina, a first-year teacher in Rust's (1994) study. Nina wrote, "I've been somewhat embarrassed that I originally wanted to teach because I loved kids. I felt that it was lacking. Now I see that actually, that's the perfect reason to decide to teach" (p. 211).

Two descriptive metaphors, a gardener and a provider, in Cole's (1990) study of beginning teachers' personal theories of teaching further attest to the centrality of nurturing in teaching, as does the description of the Norwegian "good teacher" as caregiver as well as interpreter of texts (Gudmundsdottir & Saabar, 1991, p. 5). One might speculate that teachers who are able to relate effectively to students have a measure of the tact of teaching (van Manen, 1991). They interact with children in a loving, hopeful, trusting, and responsible manner.

Table 1
Q-Group 1: Teaching Tasks of Greatest and Least Preparedness

Greatest Preparedness	*Z-scores*
Integrating learnings from two or more subject areas	2.20
Establishing positive rapport with students	1.93
Drawing subject matter from my own knowledge	1.90
Designing activities based on a collection of materials	1.39
Enriching instruction with additional content	1.29
Focusing students' attention	1.17
Least Preparedness	
Determining student grades	–2.19
Keeping daily individual student records	–1.74
Designing appropriate assessment tasks	–1.70
Evaluating appropriateness of controversial materials	–1.70
Developing appropriate means of holding students accountable	–1.51
Implementing routines to minimize time loss	–1.30
Evaluating the appropriateness of materials according to curriculum guidelines	–1.24
Monitoring the entire class while working with a part	–1.21
Understanding the underlying causes of student misbehavior	–1.01
Making smooth transitions between activities	–1.00

Group 2. This group consisted of two women and two men who, as shown in Table 2, identified themselves according to their strength of preparedness for a somewhat different set of teaching tasks. They found themselves well prepared for some aspects of planning instruction, particularly identifying lesson objectives and planning for their attainment, and for designing review exercises. They also declared themselves well prepared for some instructional delivery tasks, specifically integrating learning from two or more subject areas and summarizing lessons. They were relatively well prepared for some of the organizational aspects of teaching, for example, holding students accountable, keeping daily achievement records, and preparing the classroom for instruction.

By contrast to the relationship and creative planning tasks for which Group 1 was most prepared, the tasks for which Group 2 found it was most prepared seem more technical and businesslike. This is in keeping with Berliner's (1988) assertion that the novice stage of teacher development is a time for "learning the objective facts and features of situations and for gaining experience" (p. 40). The task for which both Groups 1 and 2 found greatest preparedness, however, was in planning instruction that integrates learning from two or more subject areas. This may be the outcome of the joint efforts of university instructors in coursework and classroom teachers in practice to address one of the central aspects of the Year 2000, a major provincial educational reform (Ministry of Education, 1990), begun with the Sullivan Royal Commission on Education and designed with extensive public consultation and teacher involvement.

Table 2
Q-Group 2: Teaching Tasks of Greatest and Least Preparedness

Greatest Preparedness	*Z-scores*
Integrating learnings from two or more subject areas	2.18
Identifying lesson objectives	1.69
Summarizing a lesson	1.67
Designing review activities	1.62
Planning for the attainment of lesson objectives	1.56
Preparing the classroom setting for instruction	1.37
Developing appropriate means for holding students accountable	1.20
Keeping daily individual achievement records	1.04
Least Preparedness	
Monitoring the entire class while working with a part	–2.19
Correcting student misbehavior unobtrusively	–1.59
Understanding the underlying causes of student behavior Problems	–1.59
Enforcing classroom rules	–1.56
Managing class according to students' maturity levels	–1.48
Handling most discipline problems in the classroom	–1.42
Relating effectively to parents	–1.35
Giving appropriate feedback to student behavior	–1.19
Promoting student self-discipline	–1.18
Communicating expectations for student learning	–1.04

Table 3
Q-Group 3: Teaching Tasks of Greatest and Least Preparedness

Greatest Preparedness	*Z-scores*
Identifying lesson objectives	2.18
Stating lesson objectives clearly	1.87
Planning for the attainment of lesson objectives	1.78
Communicating expectations for student learning	1.48
Giving clear directions to students	1.23
Providing students with a rationale for learning	1.19
Integrating learnings from two or more subject areas	1.12
Rewording questions to enhance clarity	1.01
Least Preparedness	
Enforcing classroom rules	–2.15
Promoting student self discipline	–1.86
Developing ways to improve my own teaching	–1.49
Evaluating the appropriateness of controversial materials	–1.41
Maintaining good staff relationships	–1.30
Relating effectively to parents	–1.14
Monitoring the entire class while working with a part	–1.13
Estimating wait time between asking questions and choosing respondents	–1.10
Developing alternate activities to achieve the same objective	–1.05
Determining student grades	–1.00

The tasks for which Group 2 found itself not well prepared were almost exclusively from the classroom management category, some basic and others more complex, as previously described. The group also reported lack of preparedness to communicate expectations for student learning and to relate effectively to parents.

Group 3. This group also consisted of two women and two men. They identified themselves according to the degree of their preparedness for the tasks shown in Table 3.

Some of the tasks that define their strength of preparedness were planning tasks, for example, identifying, stating clearly, and planning for the achievement of objectives. Others, such as communicating expectations, providing a rationale for learning, and giving directions, were drawn from the early and preparatory phases of instruction. Although the integration task was not *the* one for which the group identified itself as most prepared, as was the case with Groups 1 and 2, Group 3 did include that task among those for which it was most prepared.

Patterns among the tasks for which Group 3 was not well prepared were complex. Some were classroom management tasks, again from fairly routine to complex. Others focused on relationships with staff and parents. Single tasks for which preparedness was declared weak were from three areas: questioning, assessment, and planning. Rust (1994) found beginning teachers experienced difficulty both in classroom management, as did all three Q-groups, and in relating to parents as did Q-groups 2 and 3.

Q-Group summary. These three groups can be defined in terms of similarities and differences in the tasks for which they identified themselves as most and least prepared. As shown in Figure 1, only one task, integrating learning from two or more subject areas, appeared in the sets of greatest preparedness tasks for all three Q-groups. Apart from two planning tasks common to Groups 2 and 3, namely identifying lesson objectives and planning for their attainment, there were no other tasks common to any two Q-groups.

There are some elements of uniqueness in the groups. Only Group 1 felt well prepared for the more creative aspects of planning and instruction and for relating effectively to children. Group 2, with strength of preparation in more businesslike and technical tasks, was unique in identifying strength of preparedness in assessment and in a selection of tasks from across the spectrum of interactive teaching. Group 3 was distinct in declaring greater preparedness for a narrower set of tasks focused in the initial aspects of teaching, particularly preactive planning and the preliminary steps in delivering instruction.

Among the tasks for which all three groups found themselves least prepared, as shown in Figure 2, only one task, monitoring the whole class while working with a part of it, was common. Groups 1 and 2 both identified another problem task, understanding the underlying causes of student misbehavior, whereas groups 1 and 3 both expressed difficulty in evaluating

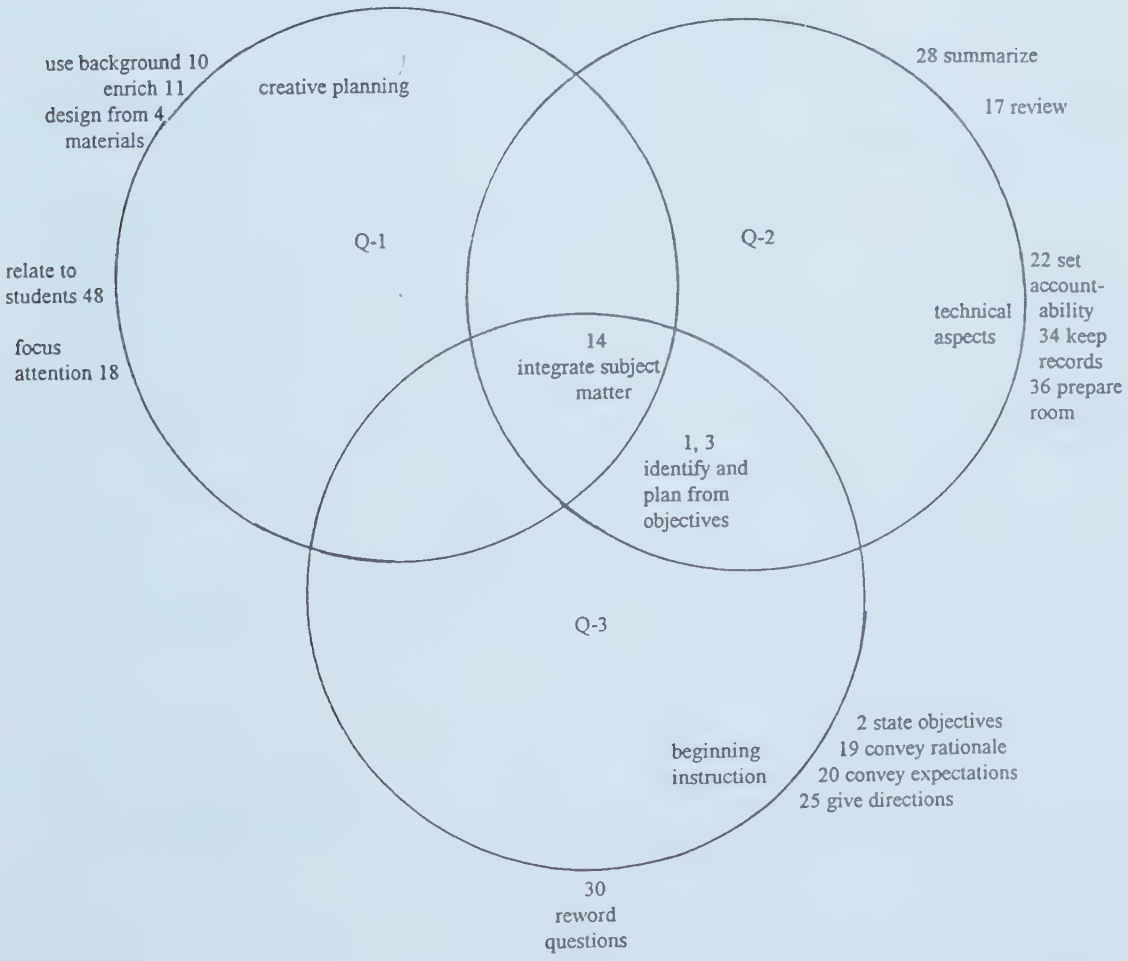


Figure 1. Common and unique tasks of greatest preparedness for three Q-groups.

controversial materials and determining grades. Groups 2 and 3 expressed common difficulty enforcing rules, promoting student self-discipline, and relating to parents. As for distinctiveness based on areas of expressed weakness in preparation, only Group 1 identified assessment as a major problem area, whereas Group 2 expressed greatest difficulty with classroom management. Group 3 included relating to staff and developing means of improving their own teaching in the wider variety of tasks for which it was less well prepared.

If one area of relative weakness in preparedness to teach were to be cited among the three groups, clearly it would be classroom management. Over 50% of the tasks in the three least preparedness sets were from this area. Despite the likely emphasis on management during the extended practicum and some preparation in university coursework, it continued to be a problem for many of the teachers. Covert, Williams, and Kennedy (1991) found that the most pressing professional development needs of beginning Newfoundland teachers were also management focused. McEvoy and Morehead (1987) drew the same conclusion regarding American beginning teachers. Indeed, Hollingsworth (1990) reported that management strategies approached automaticity only in the second or third year of teaching.

Interestingly, some of the participants in this study who continued to substitute teach or had been substitute teachers improved their management strategies in that context. One declared, "Oh, I learned more in my first month

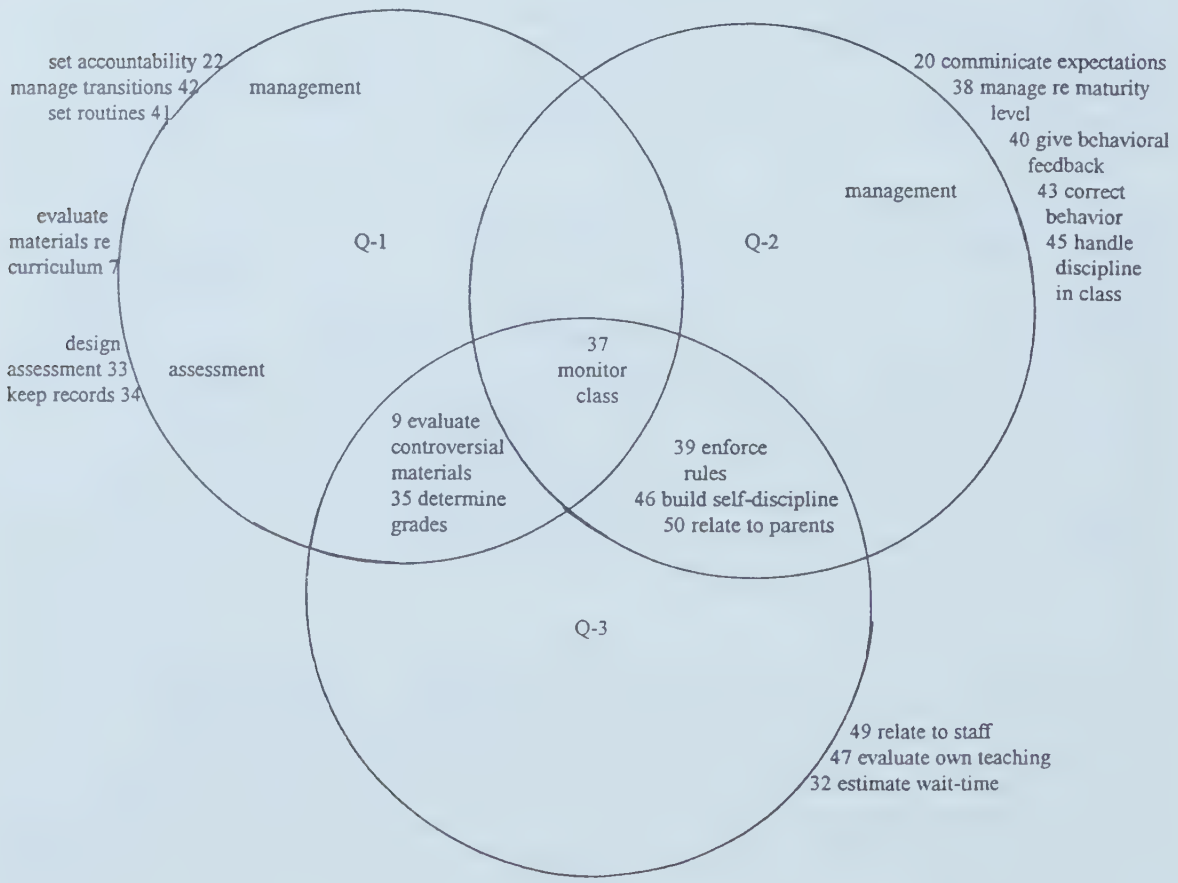


Figure 2. Common and unique tasks of least preparedness for three Q-groups.

of subbing than I did in most of the two years, as far as classroom management. As a sub, you're fair game for the kids. I would say my classroom management is improved, you know, 300%!"

Interview Findings

The structured interview (see Appendix) conducted with each of the teachers between April 28 and June 13, 1992 after the Q-sort had been returned, included 17 questions in three domains, strengths and weaknesses of preparedness to teach and subsequent effects thereof, the experience of first-year teaching, and a retrospective examination of the teacher education program as preparation for teaching.

Strengths and weakness in preparedness to teach. Because interviews were based on the tasks actually placed in the top and bottom two categories of preparedness by each individual, as contrasted with Q-group scores that appear in prior tables, comments in this section pertain to total group reports of preparedness to teach and related effects thereof. As a group, the beginning teachers in this study found they were best prepared for two tasks, integrating learning from two or more subject areas and developing good rapport with children.

Teachers' comments regarding integration, the reinforcement of concepts integral to one subject through their relationship to concepts from other subjects by various means including integrated courses, school wide foci, or community outreach were relatively few, and mainly that it had been practiced in classes and practica. One teacher observed that the integration task was not as difficult as some of the more technical aspects of teaching. Another saw it as a basic characterization of her approach to teaching because it was expected by her pupils, whom she described as creative. It was central in the work of the teacher of handicapped students, who found "no magical textbooks" and soon came to realize that her success depended on her own creative approach to teaching, including her integrative skills.

Establishing rapport with children is something many elementary teachers are able to do well and from which they derive major satisfaction. Elbaz (1992) describes the difficulty educators uneasy about the status of the profession seem to have in acknowledging the centrality of caring in teaching, when child care is often viewed in the wider society as an activity for which no special training is required. She, on the other hand, asserts that attentiveness, caring for difference, and hope comprise the moral voice in teaching, which has been largely ignored by researchers, the tendency being instead to concentrate on aspects of teacher thinking. Some of the teachers' comments on rapport with students were definitional: "more personality than anything ... the kids knowing they can trust you," or "just an ability to listen to their concerns," or "usually being able to respond appropriately; being fairly in tune or being positive." Good rapport with students seemed a valuable asset whether it "came naturally" or was carefully cultivated, whereas integration skills appeared to be an outcome of practice and sometimes the source of considerable pride.

Forty-four percent of the tasks placed in the category of second greatest preparedness focused on planning, and close to 23% on delivery of instruction, together two thirds of the total. For some, planning was rooted in the teacher

education program. One said, "all through the program, I was a meticulous planner," while another stated more directly, "they taught us how to do it, and that was a strength." As students they were enrolled in curriculum and instruction courses in all the elementary school subject areas, and, in addition, during the generic teaching skills course they were required to plan a series of lessons that would comprise if not a complete, at least a partial unit of instruction. It is not surprising, therefore, that these experiences, coupled with a 13-week practicum, in the latter phases of which they were expected to do all the teaching, would build planning strength. It appears that university courses and school experience combined to accomplish this outcome. Faced with teaching seven different classes daily, another teacher attributed being organized to "a little bit of lesson planning."

As a group, these beginning teachers reported being least prepared for classroom management tasks (44% of the tasks included two lowest preparation categories were from this area). No teacher understated the importance of management. In a confident manner, one teacher, who did not identify management as an area of weakness in preparation, speculated on the roots of his success, "My size, my voice, being male, just helps me immensely. I just have to give the evil eye and that can realign them!" With less confidence, another reported, "I'm walking a very fine line between establishing a positive rapport and having effective discipline in the classroom," while another could readily see that her growing understanding of the underlying causes of behavior was enabling her to develop genuine compassion for her students. For the total group, there was continued vigilance and some optimism with respect to management.

Another area of difficulty that gave rise to numerous comments was assessment, which was also one of the topics beginning Newfoundland and American teachers highlighted as needing attention (Covert et al., 1991; McEvoy & Morehead, 1987). Some primary teachers appeared to reject the task of determining grades, stating that they would not be used in the framework of the previously described provincial Year 2000 reform. Said one, "I thought, well, I'm teaching primary, so I don't have to assign grades per se. Now I'm teaching a 3/4 split and, in this school, grade 4s get A,B,C, and D, and I have real a skills mastery program I work very hard on! I can tell you what each child can do and to what degree and how I give that an A and that a B." Her experience prompted the statement, "That was the hardest thing. Nobody at the university taught us how to do that." Another described the incongruity between her university preparation for assessment and reporting and her current school experience as follows, "It was all standardized tests and all sorts of stuff like this, but people are throwing tests away by the carton." She complained further that students "never got an opportunity to look at different kinds of assessment techniques or practice writing report cards." In one setting beginning teachers arranged a full afternoon of consultation on assessment for themselves in an effort to acquire a broader repertoire of skills.

First-year teaching. First-year teaching is explored in this article on the basis of a subset of the questions in the Appendix, questions 1, 9, 3, 13, 16, 14, 15, 12, and 17.

Beginning a new teaching position, one has expectations about the job. Asked about the way their various experiences of first-year teaching met or did not meet their expectations and further what pleased, disappointed, or surprised them, these teachers reported many satisfactions as well as some problems.

Eight of 16 teachers had an employment history including substitute teaching. The four who continued to work as substitute teachers had hoped to find full-time teaching positions, so a basic expectation remained unmet. Still, they continued to pursue that goal, all the while learning as much as possible in the substitute's role, as had others in the study who had subsequently secured positions. Management skills, as mentioned earlier, were frequently augmented, as were instructional skills, while substitute teaching in various classrooms and schools. For example, one teacher said, "It doesn't scare me anymore, because I can do anything. I've taught French, band, music, art, computers, with or without a day plan, with or without a set lesson plan. I don't read music. And, with no day plan, I've taught music for six days as the music specialist in a school." Finding an ordered set of plans on entering classrooms was, however, the more common experience.

Successful substitute teaching is a recognized route to full-time teaching, so these teachers worked hard to be recognized as competent professionals. In one instance, this process led to a position much akin to substituting, a 70% position teaching every regular teacher's class during his or her preparation period; nevertheless, it was a more secure, and more clearly defined employment than being "half a day at two schools and a day at opposite ends of the city."

From the experiences of beginning teachers in regular positions, a pattern emerged. Initially, the euphoria of finding oneself employed under present-day conditions is considerable. "Wow, I'm finally a teacher!" This was tempered by frustrations of different sorts for different people. One teacher vividly described her entry to the classroom, as follows, "My principal just sort of dropped me off and said 'Here you go,' and I looked around. The walls were blank and I didn't like the way the classroom was.... Everything was so squished, and it wasn't my space." She went on to recall her unfamiliarity with the routines and being told nothing by anyone, rather having to ask everything, a contrast to many others who enjoyed generous staff support. At best, induction needs were sporadically met. Frustrations of locating or ordering materials, which also parallels the findings of the Newfoundland study (Covert et al., 1991), became abundantly clear when a teacher recalled saying, "Just a minute. This is a 44-page computer printout, and I have to go through this and tell you what I'm going to need next year? I don't even have a clue what I've got yet!"

Many beginning teachers confirmed their expectations that teaching is hard work, "a draining job." One drew a comparison with the 13-week practicum, saying, "Well now ... you have to psych yourself up for a whole year, and it requires a lot more mental and physical energy to really keep yourself motivated and enthusiastic for that period of time." Several mentioned the limits put on social life by having to work evenings and weekends to keep ahead of students. A few experienced a low ebb around the time of first reports, "Wow, you know this is really a lot of work and, how am I going to get through

the rest of the year?" In response, some met with other classmates every few weeks; others were assisted by staff members, whether they were appointed or self-chosen mentors or helpful colleagues. Gradually the teacher's role became a better fit, prompting the observations, "I *can* do three things at once now!" and "It wasn't as big a nightmare as I thought it would be."

The greatest satisfaction in first-year teaching was working with children; savoring their excitement as they "caught on," seeing them grow and learn to solve their own problems, garnering their respect and cooperation, making them happy, having fun with them, seeing through their eyes, and looking for the smiles on their faces. One teacher summed it up, "You come to school every day, and you know that they really want to learn.... It really inspires me to teach." A little progress on the part of a handicapped learner or a breakthrough with already jaded students were also causes for rejoicing. The enthusiasm and hopefulness of these beginning teachers were apparent in their comments, as were the hope, attentiveness, and caring for difference identified by Elbaz (1992) as evidence of the moral dimension of teaching.

The teachers were also asked, on the basis of their experience, to list what they considered the three most important skills or qualities of a first-year teacher. Classroom management and organizational skills including long and short-term planning, record keeping, and reporting progress were each mentioned by half the teachers. One remarked, "I found it frustrating because there were so many little things to keep track of, and you know, all these little systems." Five teachers included being flexible in their answers. "Go with the flow and turn on a dime," one explained. Also mentioned were resourcefulness, patience, enthusiasm, and interpersonal and communication skills.

The concerns of these teachers, apart from the ever-present one of job security, were as varied as their work situations. Most frequently mentioned was meeting the expectations of the child's next teacher: sending along a learner who is ready for the grade level. They also worried about the possible effects of labeling students, the possibility of serious student misbehavior, "fear of the unexpected," and problems of establishing rapport with parents being the person who must answer their criticisms and try to solve the problems identified; that is, "you're the bottom line." Issues of professionalism and staff politics baffled one who found it "worse than the business world." Being the "new teacher on the block" was a continuing strain for two, one of whom believed there was a constant monitoring of her state of well-being: "Will she make it? How is she standing up under the stress?"

Help-seeking was accomplished in several ways, most frequently through unofficial, self-selected staff mentors (half the group confirmed this pattern). Varying degrees of staff support were reported. Four persons stated they had no help whatsoever, whereas, in addition to the regular series of professional development programs, beginning teachers in one district reported a helpful orientation and dinner for new teachers. A few were allowed time to visit other classrooms. Still, the experiences of these beginning teachers fail to evidence organized, ongoing professional education during first-year teaching, the dawning of the 1990s "decade of induction" (Covert et al., 1991, p. 3). The extreme richness of the support systems available to some teachers is reflected in the statement of one who enjoyed, in addition to a broad base of staff

support and a close self-chosen mentor, a system of family mentoring. She said, "My mother's a teacher, my father, my husband, his mother, my aunts and uncles, my grandmother ... I was not going to be a teacher, but ..." Who could ask for more?

Asked what changes they would make in their own teaching the next year, many in the group hoped to raise the quality of instruction through better preparation, richer resources, and improved classroom management, and some even through a differently organized classroom space. Others intended to be more effective in meeting individual needs, improving the quality of assignments, or introducing more effective accountability procedures. Interestingly, a teacher who characterized himself as being able to plan successfully in an informal fashion, "by the seat of my pants," and divorced himself from formal planning, said, "I've been barely ahead of the eight-ball all year" and set as his goal "advanced planning."

The concepts of image and metaphor have been employed recently (Bullough, Knowles, & Crow, 1991; Cole, 1990; Johnston, 1992) in an attempt to move beyond the representation of teaching as an accumulation of successfully performed tasks and to capture something of its complexity. As a conclusion to the interview, participants were asked to characterize themselves metaphorically, "As beginning teacher, I'm a ____." Three themes, growth, survival, and performing tasks simultaneously were represented in the completions of the open-ended statement. The completions reflecting growth were " a little caterpillar about to be a butterfly," "a baby with much to learn and explore," "a flower ready to bloom," and simply "a learner." The survival responses were, "a survivor," "a packrat," and "a resourceful teacher." The many tasks involved in teaching prompted responses such as "a jack of all trades," "an octopus with eight arms," "a busy bee," and "a social worker cum lunch monitor cum teacher." One teacher pronounced himself simply "a committed educator," another "an entertainer," and still another "a canoe on an ocean, isolated, alone, floating." Together the metaphors comprise a rich characterization of first-year teaching.

The teacher education program as preparation. In this section, questions 2, 8, 10, 11, 7, and 6 are discussed in turn. When asked to contrast first-year teaching with the experience of practicum, teachers credited the September practicum with preparing them to begin a school year. Practicum experiences helped them plan instruction, gain acquaintance with specific content and methods, and manage student behavior. Few teachers remained in their practicum schools. The change of locale, therefore, frequently resulted in altered socioeconomic conditions, and sometimes cultural differences, both of which required adaptation. Although there was a newly found freedom in owning one's own class and not being so closely monitored, this was tempered with being "the new teacher on the block," and the full pressure of responsibility for the progress of individual students was recognized as a dimension of the more intense relationship with parents.

Asked to describe a teacher education assignment that stood out as very worthwhile, most of the teachers cited practical, immediately applicable ones that became rich resources, for example, a science unit on ice cubes, a project on fairy tales or a mathematics unit on integers, each complete with information

about where to obtain additional resources. One teacher valued having learned to play an instrument, whereas another appreciated having acquired skill in oral presentation. Two others reported having completed projects that produced deeper understandings of children from varied backgrounds or children at one or another stage of development. Having an opportunity to tutor an individual child was also valued highly. Teachers felt that the skills learned as well as the products created in the form of units, lessons, or plans were immediately useful.

One way to capture an individual's impression of the importance of features or qualities of an experience of which he or she has been a part is to ask how that person would plan a similar experience for others. So beginning teachers were asked the qualities they would consider important in the partnership and the experiences they would emphasize for a student teacher were they to become school advisors or sponsor teachers. They described the qualities of any good relationship: trust, flexibility, open communication, and clearly focused expectations, all of which would be desirable, they believed, in the student teacher's relationship with his or her school advisor. Some recalled tense relationships with cooperating teachers as documented elsewhere (Friesen-Poirier, 1992; Guillaume & Rudney, 1993; Lasley & Applegate, 1985).

Experiences such as developing one's own teaching style and progressing individually toward full responsibility for teaching, which they chose as important to emphasize in working with a student teacher, underscored the existence of individual differences among student teachers (Guillaume & Rudney, 1993; Powell, 1992) and cast some doubt on the adequacy of an apprenticeship mode of teacher education[1] where a student apprentice is expected to assume the teaching style modeled, which may or may not be appropriate. In advocating student teachers engage in all facets of teaching and participate in extracurricular activities, they further emphasized the need for varied preparation. They also stressed the importance of developing good management skills, organizing effectively, and improving instructional techniques. They looked on planning in partnership, keeping a journal, and observing other teachers as avenues to learning more about teaching.

Similarly, they were asked to imagine a course they might design and offer in the preservice teacher education program, describing its emphases and giving it a title. Some emphasized resource acquisition, for example, "Building Your Materials" or "Units in a Bag"; others would develop survival guides, for example, "Survival Tips for Subs" or "Surviving First-Year Teaching." Overviews of effective teaching were suggested, as was an exposé of possible pitfalls, "A Year in Teaching: The Truth." Other possible offerings focused on time management, the rhythmic cycles of school time, parenting skills, multiculturalism, and the expectations held for staff members. Without exception their suggestions were practical.

Asked if there were problems or issues not sufficiently addressed in the teacher education program, the teachers raised three: beginning the school year, accessing accurate information about employment prospects, and ensuring exposure to a variety of teaching models and styles. The latter is necessary if teachers construct their expertise on the basis of case-by-case response to the

problems they encounter in a "situated cognition" mode discussed by Lampert and Clark (Sparks-Langer & Colton, 1991).

Responses to a request for two or three important suggestions to improve the teacher education program varied. Some highlighted the importance of teaching experience in the backgrounds of instructors and faculty advisors, hence the importance of including classroom teachers in the instructional team. Others dealt with specific program features, for example, preserving and lengthening the September practicum to a full term. It was frequently suggested that practical instruction be increased and theoretical content reduced. In addition, some specific inclusions were proposed: report card writing, skills for substitute teaching, and more coverage of classroom management, a precise mirror of the findings of the earlier mentioned study of the professional development needs of Newfoundland first-year teachers (Covert et al., 1991) and also in keeping with Rust's (1994) description of the unmet expectations of beginning teachers.

Conclusion

The teachers in this study analyzed their preparedness to teach revealing three patterns of strength of preparedness (Q-groups). They also described the satisfactions and problems of first-year teaching and revisited their teacher education program, making suggestions for improvements. Conclusions are presented in three sections corresponding to the three areas explored in the interviews. The first section also includes conclusions relating to the Q-group findings.

Establishing Basic Skills

First, one might conclude on the basis of tasks for which the three Q-groups found themselves well prepared that these beginning teachers conform to a considerable degree to Berliner's (1988) description of the novice teacher as someone whose attention is necessarily directed to laying the groundwork for professional decision making through learning the basic strategies for classroom management and instruction. Reynolds (1992) made the same point in describing the difficulties of beginning teachers as a function of an understandable "lack of well-developed instructional routines and a meagre understanding of content-specific pedagogy" (p. 24). Near the end of a year of teaching preceded by a 13-week practicum, these teachers expressed greatest confidence in their abilities to plan and prepare for instruction and to perform a limited number of routine instructional tasks. Their abilities to reflect on various instructional strategies, weigh contextual variables, and choose the best procedures to deal with more complex instructional or management problems had not developed to the same degree. Many of their least preparedness tasks require such deliberation; for example, managing according to the maturity level of students, developing student self-discipline, evaluating materials and one's own teaching, handling discussions, correcting misbehavior unobtrusively, and giving behavioral feedback. Expert teachers, however, also seek to improve their execution of such tasks, so it does not follow necessarily that the teacher education program at UBC should be criticized for failing to develop these complex skills. It would seem, rather, that the problems of the teacher education program at UBC as identified by these teachers' areas of least

preparedness are similar to those encountered by other beginning teachers as identified in the induction needs survey done in Newfoundland (Covert et al., 1991) and the follow-up study of graduates of Alberta teacher education programs (Miklos & Greene, 1987). Reynolds' (1992) review of the research on competent beginning teaching suggests that reasonable expectations of beginning teachers include being able to:

> Plan lessons that enable students to relate new learning to prior understanding and experiences;
> Develop rapport and personal interactions with students;
> Establish and maintain rules and routines that are fair and appropriate to students;
> Arrange the physical and social conditions in the classroom in ways that are conducive to learning and that fit the academic task;
> Represent and present subject matter in ways that enable students to relate new learnings to prior understanding and that help students develop metacognitive strategies;
> Assess student learning using a variety of measurement tools and adapt instruction according to the results; and
> Reflect on their own actions and students' responses in order to improve their teaching. (p. 12)

The participants in this study reported strength of preparation for some of these tasks such as planning and development of student rapport and weakness of preparation for others, particularly assessment and management.

Perhaps there is a case for striving to meet these basic needs in preservice teacher education and planning appropriately timed, further professional development programs to remedy the difficulties experienced by beginning teachers. The present approach to induction is uneven.

Induction: The Luck of the Draw

Learning to teach is admittedly an idiosyncratic process; different teachers have different induction needs. There is, however, enough commonality to suggest a helpful menu of services and experiences for beginners. Cole's (1990) identification of the needs for planning time, readily available resources and materials, planned staff support, peer support, and opportunity for socialization into school norms and routines are endorsed in this study, where some teachers were found to be generously supported and others were not. Some of the deficits they identified could have been addressed in the teacher education program, and the teachers were not reticent in suggesting program changes and improvements.

Believe Me! Suggestions for Teacher Educators

The first-year teachers' suggestions to teacher educators were:

1. Classroom management requires more attention. Teachers recognized the need for both information about management and experience in implementing good practice.
2. Assessment of learning including writing reports to parents also needs greater emphasis.
3. Exposure to more teaching models and various classes would be preferable to practicum experience in a single setting.

4. Information on preparation for substitute teaching, employment seeking procedures, and realistic job prospects is required.
5. Practicality should be sought both directly in coursework and through continuing to include classroom teachers in the instructional team and seeking faculty advisors who have school experience.

What response can be made to these suggestions? Teacher educators may be loath to seek or endorse such feedback from program graduates to guide their work. They claim to know what teachers need and, moreover, they may be neither willing nor able to supply recipes for teaching, but committed instead to providing ideas on the basis of which to deliberate on decisions that must be made within the context of teaching. Both preservice and inservice teachers press for practicality. Yet teacher educators advance "research-based rigour as *the* fundamental basis of initial teacher education" (McNally, et al., 1994, p. 229); it is the prerequisite to reflection. As observed by Goodson (1994):

> To date much of the research employed in teacher training has been developed from a foundational disciplinary discourse—philosophical, psychological, historical, sociological—far removed from teachers. It has been produced by scholars writing within their own contexts and resonates with their own career concerns in a "publish or perish" environment. The audience is mainly academic peers who are addressed through scholarly journals. In the profoundest sense, the knowledge is, from the teachers' point of view, decontextualized. (p. 33)

Although program graduates' perceptions of their preparedness to teach and conclusions regarding program strengths and weaknesses are not as strong as pupil learning gains as a guide to program design and change, we cannot afford to ignore them.

Note

1. Schools now play an enlarged role in initial teacher education in the United Kingdom (McNally, Cope, Inglis, & Stronach, 1994).

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Berliner, D. (1984). The half-full glass: A review of research on teaching. In P. Hosford (Ed.), *Using what we know about teaching*. Alexandria, VA: Association of Supervision and Curriculum Development.

Berliner, D. (1988). Implications of studies of expertise in pedagogy for teacher education and evaluation. In Educational Testing Service (Ed.), *New directions for teacher assessment: Proceedings of the 1988 ETS invitational conference* (pp. 39-68). Princeton, NJ: Educational Testing Service.

Bullough, R.V., Jr., Knowles, J.G., & Crow, N.A. (1991). *Emerging as a teacher*. London, UK: Routledge.

Cole, A. (1990). Personal theories of teaching: Development in the formative years. *Alberta Journal of Educational Research, 36*, 203-222.

Covert, J., Williams, L., & Kennedy, W. (1991). Some perceived professional needs of beginning teachers in Newfoundland. *Alberta Journal of Educational Research, 37*, 3-17.

Elbaz, F. (1992). Hope, attentiveness, and caring for difference: The moral voice in teaching. *Teaching and Teacher Education, 8*, 421-432.

Feinman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In W.R. Houston (Ed.), *Handbook of research on teacher education* (pp. 212-233). New York: Macmillan.

Friesen-Poirier, C. (1992). A student teacher's voice: Reflections on power. *Journal of Education for Teaching, 18*, 85-91.

Goodlad, J. (1991). Why we need a complete redesign of teacher education. *Educational Leadership, 49*(3), 4-10.

Gudmundsdottir, S., & Saabar, N. (1991). *Cultural dimensions of the good teacher*. Paper presented at the meeting of ISATT5 (International Study Association of Teacher Thinking), University of Surrey.

Guillaume, A., & Rudney, G. (1993). Student teachers' growth toward independence: An analysis of their changing concerns. *Teaching and Teacher Education, 9*, 65-80.

Goodson, I. (1994). Studying the teacher's life and work. *Teaching and Teacher Education, 10*, 29-37.

Hollingsworth, S. (1990). *Teacher educator as researcher: An epistemological analysis of learning to teach reading*. Paper presented at the meeting of the American Educational Research Association, Boston.

Housego, B.E.J. (1990a). A comparative study of student teachers' feelings of preparedness to teach. *Alberta Journal of Educational Research, 36*, 223-239.

Housego, B.E.J. (1990b). Student teachers' feelings of preparedness to teach. *Canadian Journal of Education, 15*(1), 37-56.

Housego, B.E.J. (1992a). Monitoring student teachers' feelings of preparedness to teach and teacher efficacy in a new elementary teacher education program. *Journal of Education for Teaching, 18*, 259-272.

Housego, B.E.J. (1992b). Monitoring student teachers' feelings of preparedness to teach, personal teaching efficacy, and teaching efficacy in a new secondary teacher education program. *Alberta Journal of Educational Research, 38*, 49-64.

Johnston, S. (1992). Images: A way of understanding the practical knowledge of student teachers. *Teaching and Teacher Education, 8*, 123-136.

Lasley, T., & Applegate, J. (1985). Problems of early field experience students of teaching. *Teaching and Teacher Education, 1*, 221-227.

McEvoy, B., & Morehead, M. (1987). Teacher induction: What can a university do best? *Action in Education, 8*(4), 45-49.

McNally, J., Cope, P., Inglis, B. & Stronach, I. (1994). Current realities in the student teaching experience. A preliminary inquiry. *Teaching and Teacher Education, 10*, 219-230.

Miklos, E., & Greene, M. (1987). Assessments by teachers of their preservice preparation programs. *Alberta Journal of Educational Research, 33*, 191-205.

Ministry of Education, Province of British Columbia. (1990). *Enabling learners: Year 2000: A framework for learning*. Victoria, BC: Author.

Powell, R. (1992). The influences of prior experiences on pedagogical constructs of traditional and non-traditional preservice teachers. *Teaching and Teacher Education, 8*, 225-238.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research, 62*, 1-35.

Ruddick, a. (1989). *Maternal thinking: Towards a politics of peace*. New York: Ballantine.

Rust, F. (1994). The first year teaching: It's not what they expected. *Teaching and Teacher Education, 10*, 205-217.

Sparks-Langer, G., & Colton, A. (1991). Synthesis of research on teachers' reflective thinking. *Educational Leadership, 48*, 37-44.

van Manen, M. (1991). *The tact of teaching: The meaning of pedagogical thoughtfulness*. London, ON: Althouse Press.

Yin, R.K. (1989). *Case study research: Design and methods* (2nd ed.). Newbury Park, CA: Sage.

Zeichner, K. (1993). Traditions of practice in US preservice teacher education programs. *Teaching and Teacher Education, 9*, 1-13.

Appendix
Interview Questions

1. Has your first year of teaching been what you anticipated? Explain.
2. In what ways has your first teaching position been similar to the extended practicum? Different?
3. As a result of your experience, what do you see as the three most important skills for a first-year teacher?
4. These are the Q-sort tasks you identified as the ones for which you were most prepared. How has this affected you?
5. These are the Q-sort tasks you identified as the ones for which you were least prepared. How has this lack of preparedness affected you? How have you compensated?
6. From your perspective as a first-year teacher, make what you consider to be the two or three most important suggestions concerning improvement of the teacher education program.
7. Are there problems or issues which have arisen this year which were not addressed in your program and, to your mind, shuld have been? What are they?
8. Was there a particular assignment which stands out as having been very worthwhile in your teacher education program? What made it so?
9. Which component, activity or feature of your present teaching position gives you most satisfaction? Why?
10. If you could be a school advisor or sponsor teacher next year, what experiences would you emphasize for your student teachers?
11. If you could offer a course for preservice teachers, what would it be called? What would it emphasize?
12. What changes do you plan in your own teaching next year? Why?
13. What, apart from job security, is your biggest concern as a new teacher?
14. Is it also a concern of peers who, like you, are also beginning teaching?
15. Have you a mentor? If so, is he or she matched or self-selected, and how have you been helped by this person?
16. What kinds of help have been made available by the school or the district in which you teach? How have you utilized these resources?
17. In conclusion, is there anything you wish to add to the discussion we have had? A metaphor perhaps? As a first-year teacher, I'm a ____.

The Alberta Journal of Educational Research *Vol. XL, No. 3, September 1994, 375-392*

Book Reviews

Educational Reform Revisited: From William James to Siegfried Engelmann

War Against the School's Academic Child Abuse. S. Engleman. Portland, OR: Halcyon House.

Reviewed by *John B. Connors, Canadian Union College, College Heights, Alberta*

In the field of educational theory and practice there are many fads and recycled ideas that appear every decade or so. When I was in college in the 1960s, education classes included such popular buzzwords as the whole-word method, the new math, open education, values clarification, performance contracting, and programmed learning. In the 1990s these terms are seldom used and a new collection of buzzwords now includes such terms as whole language, reflective teaching, manipulative math, multicultural education, cooperative learning, intrinsic motivation, learner verification, and computer-assisted instruction. An astute observer may deduce that the name changes are only old wine in new bottles. What is less obvious is that all of the above have yet to be shown to be effective in accelerating students' educational performance.

Calls to reform the educational system in North America go back at least as far as a century. William James, in his *Talks to Teachers* in 1892, tried to persuade educators to use advances in psychology to implement changes in the classroom. John Dewey, in his *Democracy and Education* in 1916, tried to show the importance of educating the populace in order for a democratic society to function. James and Dewey advocated using research as evidence for what works, and both suggested ways to use the schools to provide free and equal opportunity for social advancement. These ideas are still hotly debated and unresolved issues in educational circles today.

For the last half-century, a small number of books have appeared each decade that have revolutionized the way we think about education, so much so that they have become classics. A short list might include the following titles:

Decade	*Title, Date, Author*
1940s	*How to solve it*, 1945, Polya
1950s	*The art of teaching*, 1950, Highet
	Why Johnny can't read, 1955, Flesch
	Handbook of educational objectives, 1956, Bloom
1960s	*The process of education*, 1960, Bruner
	Compulsory mis-education and the community of scholars, 1962, Goodman
	Why children fail, 1964, Holt
	Equality of educational opportunity, 1966, Coleman
	Learning to read: The great debate, 1967, Chall
	The technology of teaching, 1968, Skinner
	Freedom to learn, 1969, Rogers

1970s	*Crisis in the classroom*, 1970, Silberman
	Deschooling society, 1970, Illich
	Must we educate? 1973, Bereiter
	Why Johnny can't add, 1973, Kline
1980s	*On learning to read*, 1981, Bettelheim & Zelan
	Why Johnny still can't read, 1981, Flesch
	High school: A report on secondary education in America, 1983, Boyer
	The closing of the American mind, 1987, Bloom
	Cultural literacy: What every American needs to know, 1988, Hirsch

Although many of these books were and are controversial, they have so stimulated our traditional thinking about educational techniques and values that they are required college reading for many of today's future school administrators and teachers. For the 1990s, the most controversial book may be Siegfried Engelmann's latest polemic entitled *War Against the School's Academic Child Abuse* (1992). If the title doesn't shock you, the contents certainly will.

Engelmann has a long history of bucking the educational establishment. With a BA in philosophy (with an emphasis in logic) from the University of Illinois at Champaign/Urbana as his only academic credential, he has gone on to become one of the most influential advocates for educational reform in the last three decades. Precocious as a child, he was chosen to participate in the Whiz Kids program at the University of Chicago back in the 1950s but declined due to the long commute involved from the suburbs. After college he worked in a variety of jobs for which he had no specific training, such as leasing and drilling oil wells, managing automobile agencies, editing a science encyclopedia, and working in an advertising agency. During this last job, he had a contract with a toy company and was researching how often ads should appear on television before children would become saturated or lose interest, an area we now call stimulus salience. Finding little published research, he designed some experiments himself using preschool kids from his neighborhood. This experience led him to the conclusion that children think quite logically if they are given consistent explanations. He then devised a program to teach four-year-olds how to calculate simple algebraic equations, such as finding the area of a rectangle, and sent an audiotape of the teaching sessions to the Institute for Research on Exceptional Children at the University of Illinois. At the time in 1964, Carl Bereiter was in charge of an academic preschool for disadvantaged children. Bereiter was so impressed that he hired Engelmann as a research assistant to work on the project.

Those were the days of the beginning Head Start programs in the United States in which children would spend much time on nonverbal activities such as puzzles, pasting, and coloring. The Zeitgeist of the time was the horticultural metaphors of Jean Piaget and Arnold Gesell who felt that maturation would take place slowly and in discrete developmental stages that could not be rushed. Consequently, both Bereiter and Engelmann were surprised when they compared the language abilities of disadvantaged black preschoolers with those of gifted children of university faculty. Although both groups of children could use language to get along socially, the former group could not use language to express ideas. The result was the development of a remedial

language program that essentially taught English as a second language, avoiding word definitions per se while emphasizing many of the logical words used in following instructions such as *not, or, and, all, some, only,* and *if-then*. The sequence of teaching concepts was to juxtapose a wide range of positive examples of a concept contrasted with non-examples of concepts in order to define the concept boundary. These presentations were innovative compared with traditional teaching lessons that presented too narrow a range of positive concepts that encouraged undergeneralization, and not enough non-examples of concepts that induced overgeneralization of concepts (Markle, 1969).

In order to catch up with their more advantaged peers, it would be necessary to speed up the pace of their learning to more than double its present rate. To achieve this accelerated instructional pace, both teacher's explanations and student responding had to proceed rapidly. Scripted teaching lessons that were consistent with only one logical answer were devised and students were taught to respond in unison at least every 10 seconds. Specific correction procedures were used that made all students practice the error discrimination, even if only one student made the mistake. The group could not move forward until all students had mastered the concepts.

The book describing their program, *Teaching Disadvantaged Children in the Preschool* (Bereiter & Engelmann, 1966), was attacked on all angles by professional educators (e.g., Kanner, 1967). Compared with traditional teaching methods, Engelmann and Bereiter's tactics of using fast-paced verbal drill and repetition exercises had all the charm of an army boot camp during basic training. Educators labeled them as the two Young Turks (not an ethnic description of either; Engelmann is of German-Russian descent) who used pressure cooker methods for four-year-olds that resulted in children being subjected to stress and regimentation that resulted in destruction of creativity. Sociolinguists claimed that Bereiter and Engelmann didn't understand "black English" or know the difference between thinking and speaking (Pines, 1966). In spite of these criticisms, these students also showed an average IQ gain of 30 points in just 2 years (Engelmann, 1970). An early account of this program was published in Canada (Engelmann, 1967).

Bereiter and Engelmann then went on to develop the famous DISTAR programs (an acronym for Direct Instructional Programs for Teaching and Remediation) to teach disadvantaged children reading, language, and math skills in the early elementary grades. When Bereiter left to take a position at the University of Toronto (where he later helped develop the *Open Court* reading programs), Engelmann aligned himself with a behavioral psychologist, Wesley Becker. Together they entered DISTAR as a contestant against a dozen other educational models in *Project Follow-Through*, the largest United States government funded quasi-experimental research program in the history of education. The project was supposed to follow up on the educational initiatives with preschool children who had completed Head Start programs and were now entering elementary school. It was set up as a horse race, and the winner was supposed to serve as a model for effective practice for future federal funding. Eventually, 75,000 children from low-income families were served in 170 communities in the primary grades of K-3 at a cost of $1 billion. Results showed that the Engelmann and Becker Direct Instruction (DI) model was far superior

to all other models in basic skills, cognitive skills, and affective measures (Becker & Carnine, 1981). However, because the DI model contradicted the educational philosophy of the time and due to political considerations, funding was not pulled from the ineffective programs and the results were only reluctantly published in *The Congressional Record.*

When the University of Illinois refused to let them be involved in training teachers, Engelmann and Becker moved their research program to the University of Oregon in the late 1970s. There they were involved in teacher training for both regular and special education teachers, supervised student research in a doctoral program in special education and at the Oregon Research Institute, and published many curricular materials with their own Engelmann-Becker Corporation. Along with Carnine and others, they continue to publish results of the effectiveness of their instructional materials and have now developed a comprehensive theory of instruction that has received little notice in educational circles (Engelmann & Carnine, 1991). Engelmann et al. have since developed a basal reading and math program for grades 1-6, additional remedial programs for upper elementary and junior high students, tactile stimulation techniques to teach deaf and hard of hearing students, videodisc instructional programs to supplement high school math and physics courses, generalized compliance training for students with severe behavior problems, and teacher training workshops. The Association for Direct Instruction was founded in 1980 and publishes a quarterly newsletter, a dozen or so books, has DI special interest groups in professional organizations (e.g., The Association of Behavior Analysis), and sponsors numerous teacher training inservices in North American and around the world including Australia and New Zealand.

At the point in his career when most academicians are content to rest on their laurels, Engelmann continues to battle the educational establishment, from state boards of education to textbook publishers to professional organizations. An earlier book, *Your child can succeed!* (Engelmann, 1975), attacked the traditional teacher training programs in schools of education. Its purpose was to urge parents to advocate effective instructional methods and curricular materials for their own children instead of allowing the schools to label children as unteachable. His latest book is an attempt to describe in plain English why he is so frustrated in convincing those in power to think logically and rationally and not depend on past biases or political convenience. Besides the United States federal government, Engelmann singles out for his attack the State Board of Education in California, the International Reading Association, the National Council of Teachers of English, the National Council of Teachers of Mathematics, and most educational publishers of basal series in reading and math.

Engelmann documents his legal battles in Chapters 4-7 with the California State Board of Education who have a state-wide textbook adoption process. Engelmann's basal series, *DISTAR Reading Mastery*, did not make the short list of the Board's recommended basal texts, despite more than adequate proof of its effectiveness. The reason was that the Board had adopted the philosophy of the whole-language initiative that stresses comprehension over phonics. The California adoption process is a "charade" according to Engelmann because they don't follow their own rules and make up new ones to suit their purposes.

California's Curriculum Commission rejected Engelmann's *DISTAR Reading Mastery* series so that it could not be considered for adoption anywhere in the state. However, for the 102 comments made by the commission, "73 are factually incorrect," and most of the rest were "just plain dumb." Engelmann is incredulous that the media doesn't pick up on their flagrant inaccuracies, misinterpretation of data, and absurd criteria because state curriculum decisions have a broad impact on society.

Whole-language approaches are pushed by Kenneth Goodman, who is a past president of the International Reading Association (IRA). His view is that literary meaning will provide the context for which beginning readers can guess at words they don't know and that phonics should be learned incidentally and not systematically taught. Much energy in the book is spent on ridiculing Goodman's theories that appeal to what Engelmann eloquently describes as "brain-dead logic." An example, from one of Goodman's quotes describing the process of reading, is as follows:

> Early in our miscue research, we concluded that a story is easier to read than a page, a page easier to read than a paragraph, a paragraph easier than a sentence, a sentence easier than a word, and a word easier than a letter. Our research continues to support this conclusion and we believe it to be true. (Engleman, 1992, p. 30)

Engelmann's sarcastic refutation is as follows: "How could the kid read a sentence without being able to read the component words? (p. 30)" Unfortunately, attempts to reform the system by challenging the Board's decision making process or taking the Board to court were ineffectual. As a result the Board was free to choose any program it wanted regardless of the existence of any supporting proof that a particular curriculum actually works.

The National Council of Teachers of Mathematics (NCTM) is lambasted for recommending that problem solving skills should be taught before computational skills; for stressing the use of manipulatives rather than equations; for emphasizing exploratory experiences rather than number facts. According to Engelmann, the NCTM then had the audacity to promote a new set of math standards in 1989 without ever field testing its recommendations first. One math puzzle suggested as a learning activity by the *Standards* is as follows:

> Who am I?
> I have three or four sides.
> All my angles are equal.
> My sides are not equal. (p. 113)

Engelmann responds that there is no three-sided figure in traditional geometry that meets this criteria. The *Standards* also describes mathematics as an "ill-structured discipline" which implies that argument and debate are natural consequences of math interpretations. Finally the *Standards* first suggests that all its recommendation should be used nationally in math instruction and then says that pilot studies should be carried out to confirm its effectiveness. This is likened to the Federal Drug Administration (FDA) recommending distribution of a particular drug before it was first put through a rigorous trial of tests in a pilot study (Carnine, 1992).

Another example of educational illogic is taken from California's *Language-Arts Framework* in which the following teaching suggestions are give on how to teach spelling:

> [The students] only spell when they write, and the only words they need to know how to spell are the words needed in writing; therefore, one of the best instructional strategies is to generate students' spelling lists from their writings. (p. 43)

Engelmann cogently points out that this language-experience strategy neglects to teach words that will be needed for future writing or even on present writing in different topic areas.

Other chapters criticize textbook publishers for not field-testing their curricular materials before releasing them, colleges of education for not teaching proper instructional sequencing and correction procedures, and businesses for not recognizing what schools need to change in order to get results. Engelmann goes on to label the system in North American as "sick because the vast majority of people in it—from educational researchers to teachers—lack technical expertise of the single aspect of the schools that justifies their existence—instruction (p. 13)." Although Engelmann admits that no one agency can be specifically faulted, his vision is that true reform will come "only when informed citizens become educationally literate and place demands on the schools, [U.S.] Feds, publishers, and colleges of education to put their actions where their rhetoric is" (p. 13).

The book is readable with little of the jargon or convoluted arguments typical of many educational reform treatises. Engelmann's training in basic philosophical logic is invaluable as he tries to decipher the hidden meaning and agendas in much of educational policy and planning. He is most livid when he describes the "outrageous ... sorting-machine philosophy that has characterized U.S. public education from its inception" (p. 58). He asserts that it is incompatible with intelligent instruction. His own philosophy is quite simple: "If the kid hasn't learned, the teacher hasn't taught" (p. 62).

There are some ad hominum arguments that slip through, such as Engelmann stating that "possibly the most fascinating of arguments presented by the Goodmans and other whole-language activists is that nobody laughs at them" (p. 32). School administrators are set up as preaching a "rhetoric ... of avant-garde approaches (while) practising belch from the Model T sorting machine" (p. 67). Educational theorists are called "sophistic metaphysicists who perseverate on theory and have a gingerbread house notion of how kids should learn" (S. Engelmann, personal communication, June 6, 1994). In general, he scoffs at educators who are often not experts in teaching kids, training teachers, or organizing schools. In fact, Engelmann will challenge any of his detractors with a public debate by taking any 10 children, dividing them into two groups of five each, with each debater teaching a concept within a time limit. So far, he has had few takers.

The book ends on a somewhat optimistic note with the last three chapters devoted to how to change the system. It won't be easy but if the nation is serious about educational reform then these components are necessary: (a) assessment to permit timely identification of problems; (b) the program must support quality-control using teacher-trainers; (c) the trainer must have the

authority to give principals and teachers assignments; and (d) salary increases and promotions must be based on meeting or exceeding expectations. In Engelmann's ideal system the schools would operate as follows:

> The role of the board would be to state how student performance should change. The administration would provide honest and timely evaluations of the changes that are occurring. The public would receive honest and timely information about the status of students in the system and information about progress. (p. 175)

The focus of the reform, however, would still be the curriculum because "teachers don't teach without content" (p. 176). The only way to guarantee change occurs when the "administrators must suffer when the kids fail" (p. 181). In this sense, Engelmann blames the decision makers in education and sees the teachers as much victims as the students. He claims that nested accountability starts at the top: "If the superintendent's job is on the line, the district will tend to be much more scrupulous about finding out what works" (pp. 181-182).

The analogy is made between the superintendent and a coach for an athletic team: if the team isn't winning, the coach better try different strategies or he won't be kept on for the next season. The playing field rather than rhetoric decides the outcome.

The end of the book lists six rules of evidence the public should demand of its schools:

1) Don't install an approach unless you have substantial reason to believe that it will result in improved student performance.
2) Don't install an approach without making projections about student learning.
3) Don't install any practice without monitoring it and comparing performance in the classroom with projections.
4) Don't install an approach without having a back-up plan.
5) Don't maintain failed plans.
6) Don't blame parents, kids, or other extraneous factors if the plan fails. (pp. 182-188)

The absence of these details accounts for many of the problems in education today. Instead of global reforms that sound good but accomplish little, a specific quality-control system such as this would actually give teachers more autonomy if they could prove that what they were doing works. Engelmann ends his book with some suggested political and legal methods to make schools change. He also mentions a new children's advocacy group, the International Institute of Advocacy for School Children (but doesn't give an address), which has been formed to combat the academic child abuse, defined as "the use of practices that cause unnecessary failure of foundation skills" (p. 67).

Educational reformers have always had an uphill battle. Early reformers went more on hunches and intuition and tried the persuasion of case examples as circumstantial evidence. Today's meta-analytic studies are more scientific but no more easily accepted. William James (1958) used to say that any new idea in education went through three cycles: it ain't true, it ain't new, and we knew it all along. The modern counterpart in educational reform is that it takes three decades to achieve professional respectability: the first decade—your

results aren't valid and your assumptions are wrong; the second decade—the results may be valid but not for the reasons you claim; the third decade—the results and assumptions appear to be valid, but it's just different terms for what we've been saying all along! With any luck, Engelmann's 30 years of being cast out into the educational wilderness may finally be coming to an end.

Despite its title this is not just an angry book. It is more a plea for educators to act rationally and morally in making true reforms work. It is a call to the public to become literate in educational policies and to protest when they are abused. Perhaps when that happens, the constant recycling of educational ideas and periodic calls for reform will coalesce into a quality-control system that avoids victimization of students and is accountable to parents. Only then will education meet John Dewey's goal of becoming the vanguard of democracy. In this sense, William James' vision of knowledge learned as the means for making a better world is of such importance that educational reform is truly "the moral equivalent of war."

References

Becker, W., & Carnine, D. (1981). Direct instruction: A behavior theory model for comprehensive educational intervention with the disadvantaged. In S. Bijou & R. Ruiz (Eds.), *Behavior modification: Contributions to education* (pp. 145-210). Hillsdale, NJ: Lawrence Erlbaum.

Bettelheim, B., & Zelan, K. (1981). *On learning to read: The child's fascination with meaning*. New York: Knopf.

Bereiter, C. (1973). *Must we educate?* Englewood Cliffs, NJ: Prentice-Hall.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool*. Englewood Cliffs, NJ: Prentice-Hall.

Boyer, E. 1983). *High school: A report on secondary education in America*. New York: Harper & Row.

Bloom, A. (1987). *The closing of the American mind*. New York: Simon & Schuster.

Bloom , B. (1956). *Handbook of educational objectives: The classification of educational goals*. New York: David McKay.

Bruner, J. (1960). *The process of education*. New York: Vintage Books.

Carnine, D. (1992). The missing link in improving schools—Reforming educational leaders. *ADI News, 11*(3), 25-35.

Chall, J. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.

Coleman, J. (Ed.). (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Dewey, J. (1944). *Democracy and education*. New York: Free Press. (Originally published in 1916).

Engelmann, S. (1967). Teaching formal operations to preschool advantaged and disadvantaged children. *Ontario Journal of Educational Research, 9*, 193-208.

Engelmann, S. (1970). The effectiveness of direct instruction on IQ performance and achievement in reading and arithmetic. In J. Hellmuth (Ed.), *Disadvantaged child* (Vol 3). New York: Brunner/Mazel.

Engelmann, S. (1975). *Your child can succeed!* New York: Simon & Schuster.

Engelmann, S. (1992). *War against the school's academic child abuse*. Portland, OR: Halcyon House.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and applications* (rev. ed.). Eugene, OR: ADI Press.

Flesch, R. (1965). *Why Johnny can't read*. New York: Harper & Row. (Originally published in 1955).

Flesch, R. (1981). *Why Johnny still can't read: A new look at the scandal of our schools*. New York: Harper & Row.

Goodman, P. (1962). *Compulsory mis-education and the community of scholars*. New York: Vintage Books.

Highet, G. (1950). *The art of teaching*. New York: Vintage Books.

Hirsch, E.D., Jr. (1988). *Cultural literacy: What every American needs to know*. Boston: Houghton Mifflin.

Holt, J. (1964). *Why children fail*. New York: Pitman.

Illich, I. (1970). *Deschooling society*. New York: Harper & Row.

James, W. (1958). *Talks to teachers*. New York: Norton. (Originally published in 1892).

Kanner, L. (1967). Review of *Teaching Disadvantaged Children in the Preschool. American Journal of Orthopsychiatry, 37*, 976-977.
Kline, M. (1973). *Why Johnny can't add: The failure of the new math*. New York: Vintage Books.
Markle, S.M. (1969). *Good frames and bad* (2nd ed.). New York: Wiley.
Pines, M. (1966). *Revolution in learning: The years from birth to six*. New York: Harper & Row.
Polya, G. (1957). *How to solve it: A new aspect of mathematical method* (2nd ed). Garden City, NY: Doubleday Anchor. (Originally published in 1945).
Rogers, C. (1969). *Freedom to learn*. Columbus, OH: Merrill.
Silberman, C.E. (1970). *Crisis in the classroom: The remaking of American education*. New York: Random House.
Skinner, B.F. (1968). *The technology of teaching*. Englewood Cliffs, NJ: Prentice-Hall.

John B. Connors is an assistant professor of psychology at Canadian Union College in College Heights, Alberta, which is a private four-year university. Previously he worked as a school psychologist in the United States and received his doctorate from Western Michigan University in Kalamazoo, Michigan.

Understandable Ideas That Are Worth Understanding

The Unschooled Mind: How Children Think and How Schools Should Teach. Howard Gardner. New York: Basic Books, 1991, 303 pages.

Reviewed by *Robert Frender, University of Alberta*

When the astronauts of the former Soviet Union were the first into space, the American educational system was held accountable for deficiencies in its scientific and mathematical training. Recently, as the performances of both the American and Canadian economies have declined relative to those of our former Cold War allies and current New World Order competitors, our schools are being held accountable for failing to inculcate the skills and attitudes needed to compete in the new global economy. Not only have our educational institutions and our professional educators been under attack for most of my life, but the breadth and intensity of these critiques have also been steadily escalating. Allegedly, our schools are producing citizens who are mathematically inept and scientifically ignorant; both literally and culturally illiterate; either morally degenerate or saturated with Eurocentric values; historically and geographically unaware; spiritually and aesthetically deficient; and physically unfit.

With so many failings to be remedied, it is not surprising that a tremendous variety of solutions have been proposed. Some prescriptions involve more of what used to be done or more of what is currently being done: longer school days; longer school years; curricula that are refocused on the basics; abandonment of whole-language instruction; minimum competence tests; and standardized testing programs to monitor achievement. Other remedies require that different things be done: cooperative learning; inclusive education; educational vouchers; charter schools; "new" types of educational assessment—authentic assessment and performance assessment; nationally standardized educational goals, nationally standardized curricula to foster these goals, and nationally standardized testing programs to monitor the achievement of the goals; and reconceived teacher education programs. In Alberta both old and new ideas are also emerging: to stem the tides of moral degeneracy, the Mem-

ber of the Legislative Assembly from Red Deer South, Victor Doerksen, has proposed removing John Steinbeck's classic *Of Mice and Men* from the classroom. (That this prescription to alleviate moral impropriety might have some unintended side effects on cultural literacy was perhaps not envisioned by the honorable MLA.)

For much of his adult life, the distinguished developmental psychologist Howard Gardner has been exploring the human mind in a series of stylishly written books: *The Shattered Mind; The Quest for Mind; Art, Mind, and Brain; Frames of Mind;* and *The Mind's New Science*. In an earlier book, *To Open Minds: Chinese Clues to the Dilemma of Contemporary Education*, and in the book being reviewed here, *The Unschooled Mind*, Gardner has focused his synthesizing mind on how education can be reformed to better prepare young minds.

Although not dismissing many of the familiar and prevailing critiques, Gardner immediately captured my interest by identifying a generally unrecognized failing of our schools, one that I believe is more fundamental and more serious:

> Even when school appears to be successful, even when it elicits the performances for which it has apparently been designed, it typically fails to achieve its most important missions.... even students who have been well trained and who exhibit all the overt signs of success—faithful attendance at good schools, high grades and high test scores, accolades from their teachers—typically do not display an adequate understanding of the materials and concepts with which they have been working. (p. 3)

For Gardner there is one fundamental criterion that our schools should aspire to achieve: "an education that yields greater understanding in students" (p. 145). The understandings of which he speaks are the disciplinary understandings of the humanities, the arts, and the sciences. Gardner contends that these are humanity's most important cognitive achievements and that "it is necessary to come to know these understandings if we are to be fully human, to live in our time, to be able to understand it to the best of our abilities, and to build upon it" (p. 11). And Gardner disturbingly suggests that even after many years of formal education the powerful misconceptions and stereotypes of the unschooled mind remain unaltered for the majority of students. The principle issue he addresses in *The Unschooled Mind* is "Why do members of a species who master certain concepts and skills so readily exhibit so much difficulty in obtaining the skills and understandings that school at its best strives to provide?" (p. 19)

In the first part of his three-part presentation Gardner synthesizes developmental psychology's current understanding of how the intuitive, undisciplined understanding of the unschooled five-year-old mind emerges. Although these powerful ideas initially serve their creators well in a variety of contexts, they are ultimately limiting. However, these remarkably resilient ideas often prove resistant to all but the best-conceived educational interventions.

In the second part, which focuses on understanding educational institutions, Gardner first addresses what should be taught and how it might be taught, outlining in both instances the educational options that are available. He argues that we should endeavor to inculcate "a rich understanding of the

concepts and principles underlying bodies of knowledge" (p. 117), and that to achieve this goal we should adopt a "transformative" approach—one in which:

> the teacher serves as a coach or facilitator, trying to evoke certain qualities or understandings in the students. By posing certain problems, creating certain challenges, placing the student in certain situations, the teacher hopes to encourage the student to work out his own ideas, test them in various ways, and further his own understanding. (p. 119)

In the remainder of part two, using many engaging and accessible examples, Gardner thoroughly documents the misconceptions in mathematics and the natural sciences as well as the stereotypes and simplifications in the humanities and social sciences that formal schooling often fails to supplant with the hard-earned understandings of disciplined inquiry. Here is one tantalizing example: 70 percent of students who had completed a college course in mechanics gave the same naive answer as untrained students to the following problem—"What forces are acting on a coin that has been tossed straight up in the air and has reached the midway point of its upward trajectory?" (If you have any doubts about your own level of understanding, see page 3 of Gardner's book!)

In the last part of the book, Gardner begins by critically evaluating some currently popular solutions to our educational failings, many of which he dismisses as dead ends. Especially enlightening is his analysis of the strengths and limitations of progressive education, particularly its failure to establish standards to monitor the achievement of its well-intentioned goals. Most of the third part of the book offers illustrative projects and techniques that offer promise for accomplishing an education for understanding.

When assessing progressive education, Gardner argues that to succeed it requires:

> teachers who are well-trained, dedicated, and absorbed in their work ... parents who not only support the philosophy but are willing to defend it.... a community beyond the walls of the school that is hospitable to students who want to learn from its members and its institutions ... and a student body sufficiently motivated and responsible. (p. 195)

For society to develop educational institutions that instill genuine understandings in their students, it will require all of these and more. To attract into and retain in the teaching profession individuals who are capable of achieving high levels of understanding themselves, we will need well-conceived incentives. We will need to foresee the misconceptions held by teachers-to-be about the processes of teaching and learning and to develop effective teacher education programs that facilitate the transformation of these novices first into skilled apprentices and later into master teachers. We will need to graduate teachers who have a sophisticated understanding of the concepts they are trying to teach, of the learners they are attempting to instruct, and of the teaching and learning processes in which they and their students are engaged. And last, we will need to develop working conditions that will allow these highly competent professionals to successfully translate their ambitious ideals into effective educational programs that produce demonstrable accomplishments.

The Unschooled Mind is a masterful and motivating work that successfully integrates developmental psychology and pedagogical policy. Not only has the author identified a worthy goal for the noble enterprise of education, but he has also acknowledged the developmental and institutional constraints that will make the achievement of this goal a challenging task. There is no group of individuals concerned with education—undergraduates entering the profession, graduate students, experienced teachers, university professors, government policy makers, or interested citizens—who could not profitably read this sophisticatedly conceived yet accessibly written book. For initiates to the teaching profession, it can serve at least two useful functions: (a) to help transform misconceptions about what and how to teach; and (b) to provide an inspirational example of the insight and knowledge that can be developed by a dedicated scholar of the human mind. For experienced teachers, the book can help us better understand why the mission in which we have been engaged is so challenging and how we might rededicate ourselves to the teaching profession. I recommend *The Unschooled Mind* to readers who agree that ignorance is the most serious underlying cause of the many problems that confront humanity today and that understanding is our most promising antidote. I also recommend this book to those who believe that "those who can do, and those who understand teach." And for readers who are committed to a well-conceived excellence in education, I enthusiastically recommend *The Unschooled Mind*, which is unquestionably the product of an admirably schooled mind.

Robert Frender is an associate professor in the Department of Educational Psychology. His areas of focus are individual differences, the development of intelligence, human behavioral genetics, and the heredity-environment issue.

From Strength to Strength: Social Work Education and Aboriginal People. K. Feehan & D. Hannis (Eds.), 1993, Edmonton, AB: Grant MacEwan Community College.

Reviewed by *Patricia Hughes-Fuller*

> As we pulled into the driveway, I noticed blood was dripping from the trunk of one of the cars in the parking lot. I was certain I was entering a different world—and this one appeared hostile. (Wright, Chapter 5, 1993, p. 58)

From Strength to Strength is a book about Aboriginal social work education from the educator's point of view. As the above quote attests, this point of view sometimes reveals more about the insecurities that we may feel when confronting the specter of difference than it does about the complex realities of other cultures (the author later admits—with some embarrassment—that "a woman had shot an elk on her way to class and put it in the trunk of her car where she knew it would freeze quickly" p. 60).

In her introduction editor Kay Feehan refers to the need to develop "a culturally sensitive educational process" (p. 8). She adds that "no one can speak as an expert about another's culture. Only someone who lives the culture daily can identify what that culture means. Thus we saw ourselves as the content experts and our Native students as the cultural experts" (p. 9). She also com-

ments that "sustaining the balance between educational standards and local needs was a continuing struggle" (p. 8).

The contributors reflect a range of personal backgrounds, writing and teaching styles, and theoretical perspectives. The emphasis of much of *From Strength to Strength* is on education for individual empowerment and/or community development (frequently cited sources include Paulo Freire, Audrey McLaughlin, and Carl Rogers). Co-editor David Hannis discusses the potential that education has to either liberate or domesticate and draws a meaningful distinction between education and "schooling" (p. 49). Marianne Wright contends that a student-centered, Rogerian approach is most appropriate when addressing the needs of Native adult learners.

Others echo Feehan's concern about recognizing and respecting the cultural values of Native people. Kim Zapf questions whether or not orthodox social work models are capable of providing effective teaching strategies in a cross-cultural setting, and one of the more interesting episodes discussed in *From Strength to Strength* deals with a team-teaching project conducted at High Level, Alberta by Zapf and Pam Colorado. Colorado, herself an Aboriginal person, provides a vivid account of her efforts to develop a teaching model which incorporates traditional Tlingit principles.

The issues that Feehan et al. problematize are not new. The literature on Native education reflects a tension between First Nations' legitimate requests for a culturally sensitive educational process and their rejection of programming that is "watered down" (Carney, 1988; Martin, 1993; Paquette, 1989). However, according to Medicine (1987), "values are as different among Indian tribes as they are among the various Canadian social and ethnic groups" (p. 23). She adds that:

> Essentially, Natives have been educated to a Native mode. They must be aware of the non-Native mode so that they can make a better life for themselves and their children. (p. 24)

Still other Native educators question the assumption that "an educational problem rests on intractable cultural differences" (Urion, 1991, p. 8). Perhaps we should simply focus on learning about the cultures of the communities in which we work, while recognizing that our vision may be skewed by our own cultural "baggage." Medicine (1987) suggests that "an open mind in a sometimes closed system (as some Native communities are) will do much to ease the transition into a cross-cultural setting" (p. 26).

Feehan's reference to the struggle to balance educational standards with local needs raises other questions. Do instructors sacrifice students or standards, and is it really an either-or choice? How do educators sort out which aspects of their roles are actually about enforcing standards and which are merely "gatekeeping"? What does it mean to a middle-aged woman who for years has been an active and effective helper in her community to be denied certification because she didn't pass a social work course? Will she be supplanted by an officially sanctioned person, and if so, does this not imply a hierarchy in which academic knowledge is presented as having greater value than traditional practices?

It is also valid to ask how the authors reconcile the position that "conflict is the midwife of change" (Feehan & Hannis, p. 4; Hannis, p. 52) with the expressed aim of "trying to bring back the harmony and respect of the past" (Feehan, p. 9). While conflict is often a by-product of social change, is it necessarily—or in all situations and circumstances—a catalyst? As ethically responsible educators we must acknowledge that some groups and individuals are morally and culturally opposed to conflict. Surely it would be contrary to the spirit of emancipatory education to prescriptively impose a conflict-based process on these learners. Finally, Grossberg (1993) cautions that:

> a dialogic practice aims to allow the silenced to speak; only when absolutely necessary does it claim to speak for them. But this assumes that they are not already speaking because ... we do not hear them, perhaps because they are not speaking the right languages or not saying what we would demand of them. (p. 16)

Despite my concerns with the above, I think much of *From Strength to Strength* exemplifies candid, caring, and scrupulously self-critical teacher reflection. In general, the authors appear to have learned a lot from their students, and from the experience of teaching in Native communities. The one exception, in my view, is "A 'grand lady's' voyage into Nowhere Land" by Sophie Freud. Freud's academic credentials are impressive, but while the quotation marks tell us that "grand lady" was intended tongue-in-cheek, the article itself reveals that this Radcliffe alumnus probably should not have ventured into Nowhere Land (aka Slave Lake, Alberta) without a road map. Reading her description of the Native woman with "big brown eyes and a permanent uncertain smile on her face" (p. 134) that Freud says she "could not reach" (p. 134), I recoiled at the stereotype. Quigley (1990) points out that this way of viewing adult learners is typical of what he terms the deficit perspective. He reminds us that "resisters are not emotional cripples nor the misguided hard-to-reach ... resisters are courageous individuals who give their full allegiance to the culture and values they believe in, even in the face of great personal risk" (p. 114).

From Strength to Strength concludes with statements from three Native students (Carolyn Peacock, Barbara Beaulieu, and Priscilla Lalonde). Each describes her experiences in the social work program and her hopes for the future. Their comments lead me to think there is another book to be written, not from "the Native perspective" because there is no such (universal and homogeneous) thing, but rather from the perspectives—which may or may not be consensual—of the growing numbers of Native social workers who will have taken what they need from the educational system in order to serve their communities. In the meantime we have *From Strength to Strength*, an accessible collection that should be of more than local interest to practitioners of cross-cultural education.

References

Carney, R.J. (1988). The evolution of Indian adult education programs in Alberta. *Alberta Journal of Educational Research, 34*, 294-309.

Grossberg, L. (1993). Bringing it all back home—Pedagogy and cultural studies. In H. Giroux & P. McLaren (Eds.), *Between borders: Pedagogy and the politics of cultural studies.* New York: Routledge.

Martin, P. (1993). Considerations for Aboriginal adult education program planning. *Canadian Journal of Native Education, 20*, 168-175.
Medicine, B. (1987). Understanding the Native community. *Multiculturalism Education Journal, 5*(1), 21-26.
Paquette, J. (1989). Policy, power and purpose: Lessons from two Indian education scenarios. *Journal of Canadian Studies/Revue d'etudes canadiennes, 24*(2), 78-94.
Quigley, A. (1990). Hidden logic: Reproduction and resistance in adult literacy and adult basic education. *Adult Education Quarterly, 40*, 103-115.
Urion, C. (1991). Changing academic discourse about Native education: Using two pairs of eyes. *Canadian Journal of Native Education, 18*, 1-9.

Patricia Hughes-Fuller is currently completing a master's degree in adult and higher education. A former arts administrator and occasional free-lance writer, her interests include social and cultural issues, Native education, and workers' education.

Is Pedagogy Still a Dirty Word?

Designs for Excellence in Education: The Legacy of B.F. Skinner, Richard P. West & L.A. Hamerlynck (Eds.), 1992, CO: Sopris West, 284 pages.

Reviewed by *Frank J. Symons, Peabody College, Vanderbilt University*

Twenty-five years ago, Skinner (1969) suggested that effective, alternative methods of teaching were available, that these could be efficiently taught to teachers, and that their use would preclude reliance on traditional punitive arrangements designed to promote student motivation and, ultimately, success. Given the continuous cycles of reform efforts to improve education, one might wonder whether Skinner's suggestions were ever taken seriously, did they work, and, if so, why are they not being used more often?

In this collection of diverse essays, literature reviews, and first-person narratives reflecting on the past and speculating on the future of behavior analysis and education, West and Hamerlynck have brought together a series of papers reflecting on Skinner's charge and his legacy. The papers are written by informed people who pioneered many of the applications suggested by Skinner's work. In a real sense, the papers are the words of torchbearers who have endured the trials and tribulations in getting from theory to practice. Although not an academic book, the contributions or eight "selections" (not chapters) reflect the diverse strands of practice following from common underlying, experimentally derived behavioral principles. The authors provide an "insider's perspective" on the application of instructional technologies that make up the legacy of Skinner, including Programmed Instruction, Precision Teaching, Direct Instruction, Personalized System of Instruction, and Applied Behavior Analysis.

Edgar and Sulzbacher (Selection VI) in their review of Skinner's influences and effects in special education suggest that perhaps nowhere else but in behavior analysis does a system exist that offers a steady, unfolding stream of research guided by empirically derived, tested principles that lead to effective teaching techniques for those who provide services to persons with special

needs. One of the functions of special education, Baer (1972) has insightfully noted elsewhere, is almost the reverse of the regular education tradition, to accomplish meaningful development of the most difficult to teach students—quite often the least supported, the least resourceful, and the least curious students of our society. It is precisely because of trying to address the needs of the most difficult to teach, that we must understand teaching as completely as our behavioral sciences will allow. Baer goes on to illustrate that in centuries past, teaching could well consist primarily of pointing out what there was to be learned, explaining those parts that students did not learn simply by trying, and punishing any absence of trying. Students of special education with diverse needs require a great deal more than this. To teach them, it is imperative to understand the structure of what is being taught. For example, lessons that appear simple are quite often complex, and through analysis these lessons are broken into basic components that are the true simple lessons. Furthermore, it is these simple components that need to be taught prior to some of the later components becoming too complex. It is through a behavior analysis of instruction, more specifically of what it is the teacher is doing, that we learn how to arrange the environmental contingencies and come to understand the principles of effective instruction.

People may counter that the above argument is true for learners with special needs, but can a behavioral approach really demonstrate success in "regular" education? In the regular education tradition most students come to school already in possession of many prerequisite skills, and they quickly learn more in the primary grades. In fact, they are often described as a delight to teach—when what could be meant is that they teach themselves so readily. This is, essentially, Skinner's (1968) notion of the "Idol of the Good Student," which describes our belief that what the good student can learn, all students can learn. Thus the regular education students do not require us to understand very much about teaching. At the core of it, this notion illustrates our perhaps misguided belief that what works with the so-called good students should work with all students. However, current state of the practice documents the fact that a method that works with some students ("good" or otherwise) does not mean it will be effective for all. The notion of "Idols" further illustrates our failure to submit teaching and instruction to scientific analysis and our reliance on recurrent standard solutions packaged differently but selling the same content.

Students with disabilities, however, require us to learn all about the profession and practice of teaching, the more so because when we teach them well, they learn what we teach and are then on the road to learning how to teach themselves (Baer, 1972). The disciplines of special education and behavior analysis are advanced enough to be useful, but still require intensive research in understanding students' behavior in the classroom. Indeed, Edgar and Sulzbacher (Selection VI) ask whether the behavioral paradigm has run its course in special education. Implicit in their discussion is the notion that systems related changes are becoming increasingly important. In other words, the demonstration that "reinforcement works" is no longer, and should no longer be the central issue, but rather how do we get systems (e.g., schools) to be responsive and implement the best instructional technology available?

What of today's typical students? Are they still what Baer (1972) would consider the "easiest to teach"? Despite such promises implicit in America 2000 and other national programs aimed at improving education in the grade schools, it would seem that our colleges of education, the "teachers of teachers" have largely promoted ignorance rather than excellence in the "art" of teaching. As Vargas and Vargas (Selection I) suggest, it is newness, not effectiveness that gains attention for the teacher willing to try out new procedures. Under such contingencies, behavior change (i.e., what students learned) is not evaluated or analyzed; rather, teacher and student reports of how well they liked the "new technology" become the yardstick both in the grade school and beyond. In advanced education, as another example, it remains a mystery (pedagogically) why we continue to rely on a standard lecture format to train teachers. Transmitting knowledge or imparting information via a lecture that requires students to be passive recipients is a practice, Lloyd and Lloyd (Selection IV) remind us, that is antithetical to any behavior analytic understanding of what learning is. The student as an active participant is a long held fundamental tenet of behaviorism and its applications to education. Several of the contributing authors point this out in a variety of ways including West and Young (Selection III) in their discussion of Precision Teaching. They note that active student responding and a direct analysis of what the student and teacher are doing forms the cornerstone of effective instruction.

What of regular education or higher education for that matter? Are we beyond the rhetoric in designing educational programs that promote excellence (or ignorance)? It may be that applied behavior analysis is always educational, but education is not always behavior analytic. Again, Baer (1972) noted that in the past education has not had to be, because until recently, education has served the best supported, the most resourceful, the most curious students—in short, those most proficient at teaching themselves once made aware of what there was to learn. Thus our educational traditions or practices are based on a long history of interaction with the easiest to teach students. If, however, we were to regard good teachers and good students as special cases of the whole we might avoid the mistake of believing that personal experience in the classroom is the primary source of pedagogical wisdom. Skinner (1968) observed some time ago that it is difficult for teachers to profit from experience. The unfortunate solution to this, Greer (Selection VII) notes, is to try to improve pedagogical practice by introducing more sophisicated hardware such as TVs, VCRs, and computers into the classroom. Such devices, of course, will only compound the problems associated with inadequate teaching practices. One solution would seem to require that we steep potential teachers in methods of effective instruction, not through poorly applied technology, but rather through deliberate practice. The notion that in order to become a good (i.e., pedagogically competent) teacher you will need a minimum of 20 half-courses and four practicum placements over two years belies our mistaken traditions of teacher training. Consider some recent data reporting the role of deliberate practice in the acquisition of expert performance (Ericsson, Krampe, & Tesch-Romer, 1993), which documents that to become an expert requires deliberate practice in your domain for approximately 10 years eventually stabilizing at 20 hours of deliberate practice per week. How many hours of deliberate practice

has the typical, or even the best, teacher in training had by the time he or she completes a degree? Greer (Selection VII) concludes that the development of teachers who are experts in pedagogy and behavior analysis, although not sufficient, is a part of the necessary step to improving on teacher training in particular and education in general.

Twenty five year ago, Skinner (1968) wrote that "the modern classroom does not ... offer much evidence that research in the field of learning has been respected or used" (p. 19). A similar question could still be asked today. Pessimism about the current state of affairs aside, West and Hamerlynck bring together a volume that presents an interesting historical view of how Skinner and his view of behavior continues to influence educational thinking and practice. This book will introduce the unitiated and remind the experienced that a behavior analytic approach to education has a rich history of demonstration and discovery.

References

Baer, D.M. (1972). Foreword. In B. Sulzer & G.R. Mayer (Eds.), *Behavior modification procedures for school personnel* (pp. v-ix). Hinsdale, IL: Dryden Press.

Ericsson, K.A., Krampe, R.T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquistion of expert performance. *Psychological Review, 100*, 363-406.

Skinner, B.F. (1968). *The technology of teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B.F. (1969). Contingency managment in the classroom. *Education, 90*, 93-100.

Frank Symons is a doctoral student in the Department of Special Education and a research assistant at the John F. Kennedy Center for Research on Human Development. Current research interests include the assessment and treatment of severe problem behaviors in children and adults with and without developmental disabilities.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes;* any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
4-116 Education North
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1436

40

ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XL NUMBER 4 DECEMBER 1994

The Alberta Journal of Educational Research

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
University of Alberta

George Buck
University of Alberta

James Chapman
Massey University

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Linda LaRocque
University of Alberta

Stewart J.H. McCann
University College of Cape Breton

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Sharon Robertson
University of Calgary

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Graham Fishburne
Robert H. Short
Fern Snart

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, the Alberta Advisory Committee for Educational Studies, and the Research Grants Office of the University of Alberta

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Office of the Dean
845 Education Centre South
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0390

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XL, Number 4, December 1994

George H. Buck *Judy Cameron*	**393**	Introduction: Special Issue on the 40th Anniversary of the *Alberta Journal of Educational Research*
Steve Hunka	**395**	*Alberta Journal of Educational Research*: 1955-1965
H.T. Coutts *H.S. Baker*	**397**	A Study of the Written Composition of a Representative Sample of Alberta Grade Four and Grade Seven Pupils
Patricia A. McFetridge	**411**	Mathematics Learning in Young Children
Daiyo Sawada *L. Doyal Nelson*	**412**	Conservation of Length: Methodological Considerations
Erwin Miklos	**425**	Individualizing Learning
Thomas O. Maguire	**426**	Evaluation of The IPI Project
Erwin Miklos	**445**	Reflections
John McLeish	**446**	Native Education in the 1970s
A. Berger	**448**	The Education of Canadian Indians: An In-Depth Study of Nine Families
Henry W. Hodysh	**457**	Hersom and the Development of Educational Research
Naomi Hersom	**458**	Twenty-five Years of Research in Education: *The Alberta Journal of Educational Research* 1955 to 1979
Henry W. Hodysh	**472**	Educational Research in Transition: An Historical Perspective on Editorial Change
Walter H. Worth	**474**	Perceptions
Neil A. Johnson	**475**	The Pervasive, Persuasive Power of Perceptions
Walter H. Worth	**498**	Reflections

T.O. Maguire	**499**	Educational Problems Revisited
George H. Buck	**500**	M.E. LaZerte, Pioneer Educational Innovator
George H. Buck	**511**	George Murray Dunlop
Pete Kosonen *Ning Fan* *Diane J. Salter* *Carol K.K. Chan* *Gillian Bickerton*	**512**	Research Abstracts of the 1993, 1994 Dunlop Award Winners

For Sale

Back issues of the
Alberta Journal of Educational Research

$3.00 per issue

or

$10.00 for a 1-year set

Plus postage and handling charges

Some issues are available on a limited basis

Contact:

Office of the Dean
Faculty of Education
845 Education South
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
Attention: Joyce Hiller

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 393-394*

George H. Buck

and

Judy Cameron
Co-editors

Introduction: Special Issue on the 40th Anniversary of the *Alberta Journal of Educational Research*

The arrival of one's 40th birthday is usually cause for reflection and rumination. In this society, it is considered an important watershed or a midpoint in the life span. Although the anniversary of an inanimate object (such as a journal) cannot reflect upon itself, a 40th anniversary is a convenient point for others to reflect upon what the *Alberta Journal of Educational Research* (*AJER*) has contributed up to this point. Writing a summary of what has transpired, we feel, would likely alter the Zeitgeist of the *AJER* under its various editors. Thus, to present a clearer view of the journal over the past 40 years, we decided to seek out as many former editors as possible and invite them to select what they believe to be an outstanding article published in *AJER* during their tenure as editor. Former editors were also encouraged to provide a brief introduction to the article selected and/or comment on developments in educational research. Each article selected for this commemorative issue of *AJER* is reprinted in its original form (although the page numbers have been changed); editors' introductions and comments accompany the reprinted articles.

Although all the editors of the *AJER* in the past 40 years have been from the University of Alberta's Faculty of Education, there has been great diversity in their backgrounds and in their interests. It is for this reason that the *AJER* has not obtained a reputation for being the vehicle for a particular discipline or of one particular point of view. The first editor was Dr. Harold Baker, Head of the Division of Secondary Education; his successor was the second Dean of the Faculty of Education, Herbert E. Smith, who had worked with education programs at the University of Alberta since 1929. Subsequent editors include another former dean, several former department chairs, and outstanding professors who did not select administrative positions.

When we set out to invite past editors to select an article for this issue, one problem that faced us was that the first three editors (Harold Baker, Herbert Smith, and G. Eastwood) are deceased. Their tenures represent the first 12 years of the *AJER*. Given that the first years of the *AJER* are as important as subsequent years, and given our commitment to minimizing modern interpretation of the past, we decided that what was required to take the place of the first three editors was an individual who was in the Faculty of Education during the mid 1950s, who was acquainted both with educational research and those individuals associated with it, and who remains closely associated with

educational research today. These requirements were met admirably by Dr. Stephen (Steve) Hunka, University Professor of the Department of Educational Psychology.

Apart from the first 12 years of publication, most subsequent years of the *AJER* are represented by the original editors. Overall, the articles selected for this issue reflect the eclectic flavor of the journal. The studies presented here include quantitative and qualitative methodologies, historical analysis, evaluation research, and integrative reviews. This special issue concludes with research abstracts from the 1993 and 1994 winners of the G.M. Dunlop awards for outstanding master's and doctoral theses.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 395-396*

Steve Hunka

Alberta Journal of Educational Research: 1955-1965

During the first decade the *AJER* published slightly over 270 articles. A trend in the nature of these articles can be identified over the 10-year period. To understand the nature of this trend requires placing the articles in the context of the developing nature of educational research, the academic program at the University of Alberta, the state of public education in Alberta including the demography of the teacher and student population, and the goals and aspirations of those who brought the *AJER* into existence.

It appears that one goal held by the promoters of the *AJER* was to provide an educational research capability and facilities that could shed light on current educational problems in Alberta, and that would be of interest to a variety of stakeholders including the academic staff in the Faculty of Education, the teachers and administrators in the field, and supporting organizations such as the provincial government, Alberta School Trustees' Association, the Alberta Teachers' Association, the Home and School Association, and individual school boards. Then, unlike today, these organizations had little capability to carry out educational research on their own. Through the aegis of the Alberta Advisory Committee on Educational Research, an unprecedented level of cooperation between the field and the Faculty of Education to promote and carry out educational research was established that remains unrivaled to this day. Significant problems were researched, at least as perceived by academic staff members particularly in educational psychology and administration, some of whom had recently joined the University of Alberta after completing their doctoral programs. Problems that were considered significant included individual differences and disparities in educational achievement in Alberta schools, teacher education including retention and recruitment, high school dropouts, promotion policies in elementary schools, and administrative leadership. During the latter period of the decade problems of a more academic nature began to appear, partly due to interests of new staff members, and partly because the *AJER* published the results of graduate student research carried out largely by individual researchers.

Selecting one article that is in some sense outstanding during the first decade of *AJER* presents difficulty without defining what *outstanding* might be, because many different criteria could be used. For example, one might select an article because of its longstanding contribution to knowledge about education, its academic and technical merit, or its contribution to the solution of a problem current on the date of publication. A complicating factor in the selection is that the styles of educational research have changed as well as the availability of knowledge provided by other disciplines from which educational research borrows many of its procedures and techniques. In addition, facilities available for conducting educational research have changed dramatically with the introduction of the computer.

The article selected in commemoration of the first decade of *AJER*, entitled "A Study of the Written Composition of a Representative Sample of Alberta Grade Four and Grade Seven Pupils" by Coutts and Baker, is characteristic of many similar articles that provided information about the individual differences of students in Alberta schools. These articles in different subject matter areas are characterized by the use of large sample sizes and statistical comparisons in achievement among students in urban, town, graded rural, and ungraded rural schools. These articles, of course, also relate indirectly to the problems of teacher education and demographic characteristics of the population in Alberta, as well as to the instructional problems that come with classes in which students show a large degree of variability in abilities and competencies.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 397-410*

A STUDY OF THE WRITTEN COMPOSITION OF A REPRESENTATIVE SAMPLE OF ALBERTA GRADE FOUR AND GRADE SEVEN PUPILS

H. T. Coutts and H. S. Baker
Faculty of Education, University of Alberta

Background and Purpose of the Study

The present study is supplementary to a large survey of the reading and language achievement of grade IV and grade VII students in Alberta.[1] For this survey, a representative sample of students was chosen in such a way that all geographic areas, types of school organization, socio-economic and ethnic groups would be included. The city sample was drawn from Edmonton and Lethbridge, the town sample from eight towns (school population 250-1,000) randomly chosen. The graded rural sample included only rural pupils attending graded schools, the ungraded sample only rural pupils attending one-room rural schools. Slightly less than 1,000 pupils were included in each of the grade IV and grade VII samples.

The general purpose of the present study was to examine, evaluate and analyze the written composition of Alberta boys and girls at the grade IV and grade VII levels. More specific purposes were:

1. The compare achievement at these levels—both as to the quality and expression of ideas, and the correctness of mechanics and usage—among the various subsamples.
2. To explore correlations between each pair of the following variables for which data were available: intelligence, rating of the quality of ideas of an original composition, rating of correctness in mechanics and usage, and the total score on the California Language Test.

Design and Procedure

Assignment

The writers believed that, if the results were to be valid, all students at each level should be asked to write on a common topic of sufficient breadth that pupils of varied backgrounds could react to it. They further believed that it should be planned to stimulate boys and girls to marshal ideas from their own experience, to organize these ideas thoughtfully, and to write a short composition in as clear and effective a manner as possible. Actual assignments were as follows:

[1]For further information concerning this survey, see "A survey of Reading Achievement in Alberta schools" in the March 1955 issue of this Journal. See also, in the present issue, the editorial page and "A Survey of the Language Achievement of Alberta School Children," pp. 39-52.

Reprinted from *The Alberta Journal of Educational Research, 1*(2), 5-18, 1955.

GRADE IV

Joe can hardly wait to get home after school. He's building a model airplane, and in a couple of days he'll have it ready to fly. Bill heads for the playground, where he finds several friends eager to play ball with him. Sue goes straight home to read, while Mary and Ellen get out on their roller skates as fast as they can.

What do you like to do after four o'clock, when school is over for the day? Tell about it in one or more paragraphs. Use the title *Fun After Four.* Things you might mention are why you like it, how you do it, and any other interesting facts about it.

Arange your ideas and sentences as smoothly as you can, planning and trying them out on page 2, for *rough work.* Then rewrite them on page 3, for *finished work.* You needn't fill all the lines, although you should tell enough to be interesting.

Use pencil for all your work.

GRADE VII

Everyone has a hobby. Some boys and girls raise stock or grain for show. Some boys prefer building models, radio sets, or machines. Girls often like to embroider, to sew, or to do leather work. Some boys and girls collect things: stamps, coins, pictures, recordings. Others prefer to paint, to play the piano. No matter what your hobby is, you must have very good reasons for liking it. Write a composition of 150-200 words under the title *Why My Hobby is Important.* You may arrange your composition in one or more paragraphs depending upon how you want to organize your ideas.

Suggested Procedure

1. Write down the ideas you want to include.
2. Make a brief outline or plan of your composition.
3. Write your composition in rough form in the space provided.
4. Check your composition, making additions and alterations for improvement and correcting spelling, grammar, punctuation, and sentence form if necessary.
5. Copy your composition carefully in ink beneath the words *FOR FINISHED WORK.*

Evaluation and Scoring

Quality. The writers were aware of the existence of the *Hudelson English Composition Scale* and of other such scales. None seemed suitable for use in Alberta under the conditions of the present study. It was therefore decided to develop a scale specifically for the assignments made to Alberta grade IV and grade VII students, and based on the compositions written through the stimulus of the assignments.

Scoring instruction were as follows:

Evaluate the papers on a five-point scale using the following criteria for guidance. Enter the value in the space provided on the record card under the heading *Ideas, Organization, Presenation.*

Value 5: The qualities suggested for 4 plus such added qualities as the happy turn of phrase, added polish, greater maturity of thought.

Value 4: Selection of Ideas.
Unity of Ideas—stays with the topic.
Paragraphing—paragraphs used when necessary.
Use of Connectives—connectives used when necessary.
Sentence organization—sentences logically organized.

Sentence Variety—reasonable variety in sentence sturcutre to give a pleasing effect.
Style—pleasing and unpretentious style.
Appearance—agreeable as to appearance.

Value 3: The same qualities as in 4 above, but only average in effectiveness. Will include papers otherwise well written but not on the topic assigned.

Value 2: Lacking in most of the qualities as in 4 above, but intelligible.

Value 1: Generally lacking in the qualities of 4 above. Is incoherent, garbled, illogical, immature, obscure.

For papers graded 4 and 5 check under "Reasons for Evaluation" the specific points of strength which led to the evaluation placed on the paper.
For papers graded 1 and 2 check under "Reasons for Evaluation" those items in which the paper was so deficient as to merit the evaluation which you gave it.
For papers graded 3 no checks under "Reasons for Evaluation" are necessary.

Forty to fifty compositions from each of the grade IV and grade VII samples were submitted to a panel of experienced judges who evaluated them independently. When there was complete agreement by all of the judges, two scales[2]—one for grade IV and one for grade VII—were prepared and mimeographed.

Using these scales, selected students registered in the Faculty of Education, University of Alberta, evaluated the quality of the compositions in both the grade IV and grade VII samples. Each paper was evaluated independently by three students, and the average evaluation (rounded to the nearest whole number) calculated. The score values thus determined were used in making the analyses and comparisons of the quality of the compositions in the study.

Mechanics and usage. Each composition was scored according to the following instructions:

In the left hand margin of the student's paper place the letter *S* for each error in spelling, the letter *P* for each error in punctuation, and the letter *U* for each error in grammar and usage. Total the spelling, punctuation, and grammar and usage errors, and place the totals in the boxes on the record card. In the box to the right of these enter the value on usage and convention by using the following conversion scale:

Value 5—no errors
Value 4—1, 2 or 3 errors
Value 3—4, 5 or 6 errors
Value 2—7, 8 or 9 errors
Value 1—more than 9 errors.

Accept any legitimate spelling of a word. Count as an error in spelling the omission of the capital from proper nouns but not the omission of the capital at the beginning of the sentence. Count also as errors the capitalization of a word which does not require such capitalization.
Only those punctuation errors should be counted which are definite breaches or which distort the meaning. Count as errors in punctuation the omission of a capital at the beginning of a sentence. Do not count as punctuation errors the comma splice since this is really related to sentence organization rather than punctuation.

[2]These scales are reproduced in the Appendix to this Journal, pages 53-61.

Count as errors in grammar and usage only those items which are unacceptable in standard informal English. Use Perrin's *Writer's Guide and Index to English* and either Webster's *Collegiate Dictionary* or the *American College Dictionary* as sources against which to check usage.

In addition a count was made of the spelling, punctuation, and usage errors on all of the papers in both the grade IV and the grade VII samples.

Correlations

The intelligence score, the total score on the California Language Test, the *quality* score, and the *mechanics* score were recorded for each student. These data were then complied and analyzed, the statistical computations being done by graduate students under the guidance of Dr. G. M. Dunlop.

Written Expression in Grade Four

Quality

Table I shows the distribution of quality scores for all pupils in the grade IV sample. Of this sample 3.8 per cent (34 students) received a score of five, 24.3 per cent a score of four, 46.4 per cent a score of three, 22.2 per cent a score of two, and 3.3 per cent a score of one. The highest mean score was obtained by the city subsample. The town, graded rural, and ungraded rural subsamples follow in that order.

TABLE I

QUALITY SCORES, MEANS, AND STANDARD DEVIATIONS FOR FOUR SUBSAMPLES OF ALBERTA GRADE FOUR PUPILS

Sample	Scores					Number of Students	Mean	Standard Deviation	Standard Error
	5	4	3	2	1				
City	19	88	152	63	6	328	3.167	0.828	0.046
Town	7	62	92	35	5	201	3.154	0.835	0.059
Graded Rural	3	37	89	50	7	186	2.887	0.819	0.060
Ungraded Rural	5	79	29	49	11	173	2.815	0.887	0.067
TOTAL	34	216	412	197	29	888	3.033	0.865	0.029

The means shown in Table I were variously paired and their significance tested by the Cochran and Cox approximate method. Table II shows these differences, together with the standard errors of the differences between pairs, the observed *t* values, and significant differences between pairs.

TABLE II

COMPARATIVE ACHIEVEMENT IN QUALITY OF WRITTEN COMPOSITION OF GRADE FOUR PUPILS EXPRESSED AS MEAN DIFFERENCES

Sample	Town	Graded Rural	Ungraded Rural	Total Sample
CITY				
Dif.	0.013	0.280	0.352	0.134
SE_D	0.075	0.076	0.082	0.054
t	0.174	3.702	4.309	2.407
Sig.		.01	.01	.05
TOWN				
Dif.		0.267	0.339	0.121
SE_D		0.089	0.090	0.066
t		3.166	3.776	1.839
Sig.		.01	.01	
GRADED RURAL				
Dif.			0.072	—0.146
SE_D			0.091	0.067
t			0.795	2.184
Sig.				.05
UNGRADED RURAL				
Dif.				—0.218
SE_D				0.074
t				2.962
Sig.				.01

The means of the city and town subsamples are significantly greater than those of the graded rural and ungraded rural subsamples at the .01 level. Other differences between pairs are not statistically significant.

The above indications may be confirmed by reference to the mean of the total sample, which is significantly smaller than that of the city subsample and significantly greater than the means of both graded rural and ungraded rural subsamples.

Mechanics and Usage

Table III shows the distribution of scores in mechanics and usage for all pupils in the grade IV sample. Of this sample 6.1 per cent (54 students) received a score of five, 37.6 per cent a score of four, 33.9 per cent a score of three, 13.9 per cent a score of two, and 8.5 per cent a score of one. The highest mean score in mechanics was shared by the city and town subsamples. The graded rural and ungraded rural subsamples follow in that order.

Table III

MECHANICS SCORES, MEANS, AND STANDARD DEVIATIONS FOR FOUR SUBSAMPLES OF ALBERTA GRADE FOUR PUPILS

Sample	Scores					Number of Students	Mean	Standard Deviation	Standard Error
	5	4	3	2	1				
City	18	134	111	45	20	328	3.259	0.990	0.054
Town	18	75	65	27	16	201	3.259	1.057	0.075
Graded Rural	12	76	58	28	12	186	3.258	1.004	0.074
Ungraded Rural	6	49	67	24	27	173	2.902	1.084	0.082
TOTAL	54	334	301	124	75	888	3.189	1.030	0.034

Table IV

COMPARATIVE ACHIEVEMENT IN MECHANICS AND USAGE OF GRADE IV PUPILS EXPRESSED AS MEAN DIFFERENCES

Sample	Town	Graded Rural	Ungraded Rural	Total Sample
CITY				
Dif.	0.000	0.001	0.357	—0.070
SE_D	0.093	0.092	0.099	0.065
t	0.000	0.010	3.600	1.081
Sig.			.01	
TOWN				
Dif.		0.001	0.357	0.070
SE_D		0.105	0.111	0.082
t		0.101	3.205	0.850
Sig.			.01	
GRADED RURAL				
Dif.			0.356	0.069
SE_D			0.111	0.082
t			3.210	0.859
Sig.			.01	
UNGRADED RURAL				
Dif.				—0.287
SE_D				0.090
t				3.203
Sig.				.01

Table IV compares the achievement of the four subsamples in terms of mean differences. The mean of the ungraded rural subsample is significantly smaller (.01) than that of the total sample and of the other subsamples.

Written Expression in Grade Seven

Quality

Table V shows the distribution of quality scores for all pupils in the grade VII sample. Of this sample 2.1 per cent (19 students) received a score of five, 19.1 per cent a score of four, 48.5 per cent a score of three, 26.3 per cent a score of two, and 3.9 per cent a score of one. The highest mean score was obtained by the city subsample. The town, graded rural, and ungraded rural subsamples follow in that order.

TABLE V

QUALITY SCORES, MEANS, AND STANDARD DEVIATIONS FOR FOUR SUBSAMPLES OF ALBERTA GRADE SEVEN PUPILS

Sample	Scores					Number of Students	Mean	Standard Deviation	Standard Error
	5	4	3	2	1				
City	9	68	162	81	10	330	2.955	0.824	0.045
Town	5	40	115	47	7	214	2.949	0.793	0.054
Graded Rural	2	31	81	49	6	169	2.846	1.032	0.079
Ungraded Rural	3	31	75	58	12	179	2.749	0.877	0.066
TOTAL	19	170	433	235	35	892	2.891	0.827	0.028

Table VI compares the achievement of the four subsamples in terms of mean differences. The means of the city and town subsamples were found to be significantly greater than the mean of the ungraded rural subsamples (.05). The latter was found to be significantly smaller than the mean of the total sample (.05).

Mechanics and usage

Table VII shows the distribution of scores in mechanics and usage for all pupils in the grade VII sample. Of this sample 3.6 per cent (32 students) received a score of five, 25.9 percent a score of four, 31.3 per cent a score of three, 18.3 per cent a score of two, and 20.9 per cent a score of one. The highest mean score was obtained by the city subsample. The town, ungraded rural, and graded rural subsamples follow in that order.

TABLE VI

COMPARATIVE ACHIEVEMENT IN QUALITY OF WRITTEN COMPOSITION OF GRADE VII PUPILS EXPRESSED AS MEAN DIFFERENCES

Sample	Town	Graded Rural	Ungraded Rural	Total Sample
CITY				
Dif.	0.006	0.109	0.206	0.064
SE_D	0.071	0.092	0.080	0.053
t	0.085	1.079	2.578	1.205
Sig.			.05	
TOWN				
Dif.		0.103	0.200	0.058
SE_D		0.096	0.085	0.061
t		1.068	2.346	0.951
Sig.			.05	
GRADED RURAL				
Dif.			0.097	—0.045
SE_D			0.103	0.089
t			0.940	0.534
Sig.				
UNGRADED RURAL				
Dif.				—0.071
SE_D				1.990
t				1.968
Sig.				.05

TABLE VII

MECHANICS SCORES, MEANS, AND STANDARD DEVIATIONS FOR FOUR SUBSAMPLES OF ALBERTA GRADE SEVEN PUPILS

Sample	Scores					Number of Students	Mean	Standard Deviation	Standard Error
	5	4	3	2	1				
City	16	95	107	53	59	330	2.867	1.156	0.064
Town	8	53	73	36	44	214	2.743	1.150	0.079
Graded Rural	3	35	54	42	35	169	2.578	1.086	0.084
Ungraded Rural	5	48	45	32	49	179	2.598	1.222	0.091
TOTAL	32	231	279	163	187	892	2.729	1.162	0.039

Table VIII compares the achievement of the four subsamples in terms of mean differences. The mean of the city subsample was found to be significantly greater than the means of the graded rural (.01) and ungraded rural (.05) subsamples.

TABLE VIII

COMPARATIVE ACHIEVEMENT IN MECHANICS AND USAGE OF GRADE VII PUPILS EXPRESSED AS MEAN DIFFERENCES

Sample	Town	Graded Rural	Ungraded Rural	Total Sample
CITY				
Dif.	0.124	0.289	0.269	0.138
SE_D	0.099	0.103	0.110	0.072
t	1.246	2.794	2.445	1.911
Sig.		.01	.05	
TOWN				
Dif.		0.165	0.145	0.014
SE_D		0.115	0.121	0.088
t		1.436	1.200	0.159
Sig.				
GRADED RURAL				
Dif.			—0.020	—0.151
SE_D			0.124	0.092
t			1.612	1.635
Sig.				
UNGRADED RURAL				
Dif.				—0.131
SE_D				0.099
t				1.312
Sig.				

Intercorrelations

Using measures of intelligence, language achievement on the California Language Test, scores on the quality of written composition and on mechanics and usage, intercorrelations were calculated as shown in Table IX.

Although none of these figures are high, all are positive. Further reference will be made to them below.

TABLE IX

INTERCORRELATIONS AMONG SCORES ON INTELLIGENCE, CALIFORNIA TEST, QUALITY OF WRITTEN COMPOSITION, MECHANICS AND USAGE

Correlated Measures	Grade IV			Grade VII		
	California Test	Quality	Mechanics	California Test	Quality	Mechanics
Intelligence	.55	.41	.33	.70	.44	.36
California Test		.46	.42		.55	.50
Quality			.26			.40

Errors in Mechanics and Usage

A count of the errors in spelling, punctuation and grammar was made for a slightly larger sample of Alberta grade IV and VII pupils (934 and 938 respectively). Figures 1 and 2 show the results of this count. Further reference to these data will be made below.

Interpretations and Conclusions

Achievement in urban and rural schools

A summary of the relative achievement of city, town, and rural pupils in written composition is provided in Table X. In general, the relationship of these scores is comparable to that in reading found by Carmichael and Rees.[3] Their analysis of possible reasons for the relationship is applicable to the present study.

Differences in intelligence may well be one factor. Reid found that mean scores for intelligence of the pupils in the Alberta sample ranked downward from city, through town and graded rural to ungraded rural subsamples.[4] The correlations between quality, mechanics, and intelligence shown in Table IX are positive, though not high.

If intelligence is a factor, it may be in some measure a function of the pupil's environment. Cultural opportunities in many rural areas are necessarily less than in towns and cities. This is especially true of the kinds of opportunity that lead to expertness in communication: the development of verbalism bears clear relationship to the breadth of social contacts and the richness and appropriateness of available printed material. While the radio is

[3]Carmichael and Rees, **op. cit.**, pp. 25-26.
[4]Reid and Conquest, **op. cit.**, pp. 45-46.

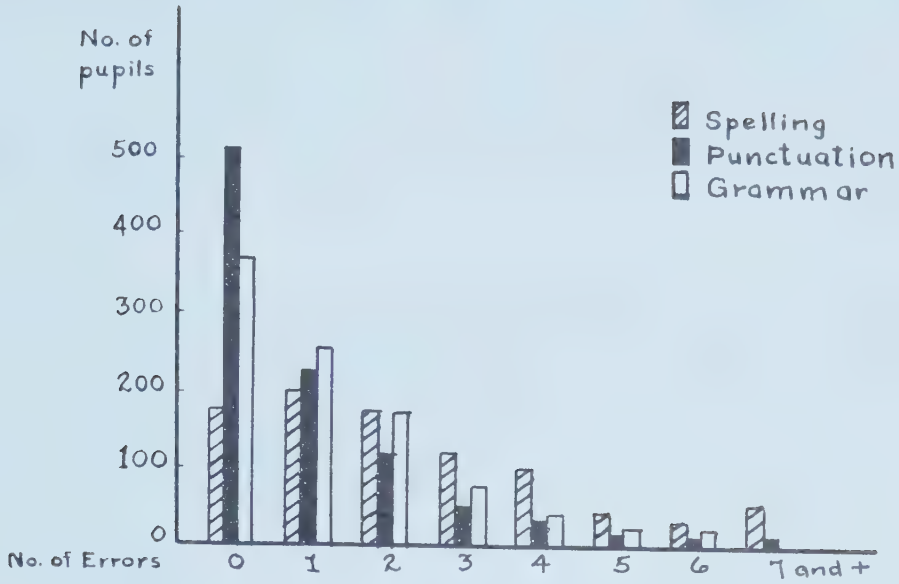


FIGURE 1. Number of errors in spelling, punctuation and grammar of a sample of 934 grade IV pupils.

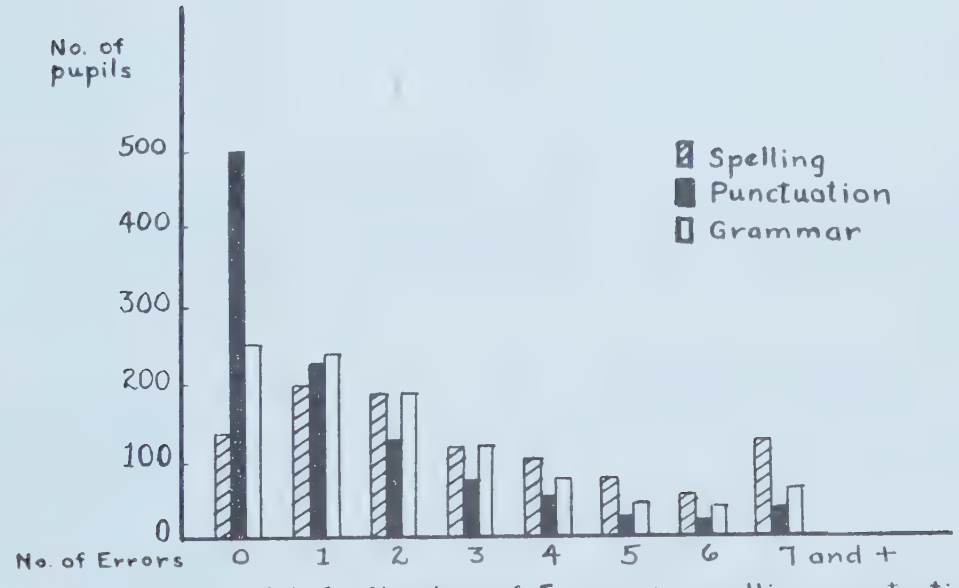


FIGURE 2. Number of Errors in spelling, punctuation and grammar of a sample of 938 grade VII pupils.

broadening the listening experiences of rural children, and while centralization and vanning are providing wider social contacts for them, it is doubtful whether their advantages are yet equal to those of town and city children.

The probability of a higher proportion of the foreign born in rural than in urban Alberta suggests a second and obvious explanation for differences in language scores. It might also suggest, in terms of the language handicap, an explanation for differences in intelligence scores.

TABLE X

SUBSAMPLES WITH MEAN SCORES SIGNIFICANTLY HIGHER THAN THOSE OF OTHER SUBSAMPLES AND OF THE TOTAL SAMPLE

Subsample	Grade IV		Grade VII	
	Quality	Mechanics	Quality	Mechanics
City	GR (.01) UGR (.01) TS (.05)			GR (.01) UGR (.05)
Town	GR (.01) UGR (.01)			
Graded Rural		UGR (.01)		

NOTE: GR—graded rural; UGR—ungraded rural; TS—total sample.
The table is to be read as follows: For grade IV, the mean *quality* score of the city subsample is significantly higher than that of graded rural and ungraded rural subsamples at the .01 level of confidence, and of the total sample at the .05 level; the mean *mechanics* score is significantly higher than that of the ungraded rural subsample at the .01 level.

The foregoing are relatively inflexible circumstances associated with the opening up of a new country. Teaching competency is—potentially at least—more controllable, and this competency may be a third factor. At any rate, it justifies conjecture with reference to differences in the scores of urban and rural students. Generally speaking, teachers with higher qualifications do seek and find employment in towns and cities. (It has been said with some truth that the rural divisions staff the city schools.) More attractive living and working conditions would encourage more superior rural teachers to remain in rural areas. An increase in teacher recruits from cities would tend to fill more of the available jobs there, further reducing the drain of competent rural teachers.

Achievement in grade IV and grade VII

The pattern of significant differences at the grade VII level is not so distinctive as at the grade IV level. There seems to be no clear reason for this finding. Perhaps three more years of school and community influence (as opposed to more exclusive and differential home contacts in the early school years) tend to level language facility.

Intercorrelations

It has already been pointed out that the correlations among scores on intelligence, the California Language Test, quality, and mechanics[5] are all positive but rather low. They are certainly too low to justify prediction or generalization with reference to the primary purposes of this study. They do, however, tend to confirm the view that achievement in mechanics and usage has less to do with intelligence than does the quality of ideas and their expression.

Spelling, punctuation, and grammar

The data presented in Figures 1 and 2 are based on an average of approximately 100 words for compositions in grade IV, and of 200 words for those in grade VII. In grade IV, in a random sample of 934 students, about 180 had no errors in spelling, 200 had only one, and 175 had two; in punctuation, some 515 of the 934 had no errors at all; in usage and grammar, 375 had no errors. In grade VII, more than 500 of the 938 students ranged from no errors to two errors in spelling, 420 had no errors in punctuation, and more than 650 had from no errors to two errors in grammar and usage.

These are encouraging figures. They suggest that large numbers of our students are writing mechanically correct English, and that their usage patterns are respectable.

Evaluating student writing

After experience with evaluation procedures in this study, much can be said in favour of a two-phase evaluation of student writing. This implies only in part the distinction between so-called "subjective" and "objective" procedures. It has more to do with the distinction between sense or spirit (quality) and form (mechanics) —a vital distinction, not only for purposes of accurate evaluation but for the much more important goals of criticism and improvement.

Evaluation for quality means the assessment of ideas and their presentation in terms of their sheer impact as communication (selection, organization, diction). Evaluation for mechanics means the assessment of formal facility or correctness (punctuation,

[5]See pages 13-14.

spelling, usage). The use of a check sheet may help in each of these processes, but it must be kept simple and uncomplicated. Otherwise the process of evaluation tends to degenerate into mere mathematical processes of addition and substraction. For the evaluation of quality, the development of local scales has been found highly desirable.

BIBLIOGRAPHY

DeBoer, J. J. and others. *Teaching Secondary English.* New York, McGraw-Hill Book Company, 1951.

Gurrey, Percival. *The Teaching of Written English.* London, Longmans Green and Company, 1954.

Hudelson, Earl. *Hudelson English Composition Scale.* Yonkers-on-Hudson, World Book Company, 1923.

Jewett, Ida A. "Simples from a Composition Reader's Scrip." *Teachers College Record* XXXIX: 1 (October, 1937), pp. 16-24.

LaBrant, Lou. *We Teach English.* New York, Harcourt Brace and Company, 1951.

Mirrielees, Lucia B. *Teaching Composition and Literature.* New York, Harcourt Brace and Company, 1952.

National Council of Teachers of English. *The English Language Arts.* New York, Appleton-Century Crofts, 1952.

Williams, G. Perrie. *The Northamptonshire Composition Scale.* London, Harrap and Company, 1933.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 411-411*

Patricia A. McFetridge
Editor, December 1967-September 1969

Mathematics Learning in Young Children

It is a pleasure to participate in this special issue of AJER featuring previously published outstanding research. I have selected an article by Sawada and Nelson for its significance as the beginning of important studies in mathematics learning in young children.

From this early study each author developed a long program of outstanding research into the nature of concept formation and mathematical understandings in young children, a field relatively unexplored at that time.

Dr. Nelson's research culminated in the Lecture Series Award from the Faculty of Education, University of Alberta, 1979. He presented a monograph entitled "Studying problem solving behavior in early childhood." Dr. Sawada's later work included a research grant from SSHRC, 1991-1994, for research entitled "Mathematics in a Literary Mode" and to mathematical resources for Division 11 in the schools entitled *Mathworld*. *Mathworld* has been nominated for the 1994 Smithsonian Award from Computerworld.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 412-424*

A modified technique for assessing the child's acquisition of length conservation was designed and tested on 62 subjects aged 5-4 to 8-0. The technique attempted to eliminate or minimize the confounding effects of dependence on verbal means of communication, no understanding of instructions, failure to perceive the initial comparison, forgetting, perceptual estimation and guessing.

DAIYO SAWADA and
L. DOYAL NELSON
The University of Alberta

Conservation of Length: Methodological Considerations

Since Piaget's discovery of conservation as a behavioural attribute involved in the child's acquisition of operational thought, many replicative studies have been done confirming the *existence* of the phenomenon of conservation with the result that conservation is generally accepted as a valid attribute in describing the child's cognitive status. These replicative investigations have given way to studies which attempt to relate the phenomenon of conservation to other cognitive variables.[1,2,3,4,5] However, the techniques used in these later investigations to obtain a measure of conservation are essentially the same as those used in the replicative studies. Unfortunately, techniques useful for showing the *existence* of a phenomenon may be neither precise enough nor powerful enough to establish the relationship of the phenomenon to other variables.

Suppose, for example, that we wished to test a simple hypothesis involving the relation of verbal and nonverbal IQ to the acquisition of conservation of length. Suppose further that we classify students as being conservers or nonconservers on the basis of the "usual" procedure typified by the following:

1. Two rods of equal length are placed in front of S as indicated in Figure 1.
 S is asked, "Which stick is longer, or are they both the same length?" Most S will say that they are the "same."

Reprinted from *The Alberta Journal of Educational Research, 14*(1), 23-35, 1968.



Figure 1.

Rods aligned in "identity position."

2. One of the sticks is transformed (e.g., pushed a few inches to the left or right).
3. S is asked, "Which stick is longer now, or are they still the same length?"
4. Sometimes the question "WHY?" is asked.

In the above procedure there is a heavy reliance on verbal means to communicate to S that he is to respond on the basis of length. In fact, the only way S has of knowing that the attribute under consideration is length is that E used the word "length" or "longer" or some such synonym. Sometimes verbal descriptions, such as "a boy walking along a path" or "two ants going on a walk", are used to communicate to S the criteria he is to use when responding. These verbal descriptions tend to encourage responses based on anthropomorphic criteria which could confound the measure obtained.

If we could couch our description of the criteria of response in a logical rather than a purely verbal form, we could not only give a clearer and more precise definition of the response criteria, but also increase the likelihood that S would respond logically, if, of course, he is capable of doing so. Braine[6,7,8] has reported studies indicating some of the disadvantages of employing a technique which is crucially dependent on verbal means of communicating the response criteria to S. Smedslund[9,10] has cast some doubt on Braine's findings, but regardless of the final outcome of the Braine-Smedslund interchange (see Gruen[11]) it still remains that the measure of conservation derived from the procedure outlined above is so heavily loaded with a built-in verbal factor that its usefulness is severely limited.

Furthermore, such a procedure, dulled as it is by an overlay of verbal dependencies, would probably not possess the necessary sensitivity for detecting the threshold emergence of conservation in young children. In designing training programs for these children it is extremely important that such threshold levels be determined with some degree of precision.

In general, what is needed is a measure of conservation which is maximally free from confounding influences. Such a measure might be called "relatively pure". This paper presents a modified technique for

assessing conservation, and reports, briefly, on an experiment using the technique as a partial evaluation of its effectiveness.

The Problem

In order to design a technique which will give us a relatively pure measure of conservation it would be beneficial to make some basic distinctions concerning the relationship of conservation to concepts in general. At a minimum, we need to distinguish between the concept of a particular attribute (e.g., number, length, time, volume), and the concept of the conservation of the magnitude of that particular attribute. Although the concept of the conservation of the magnitude of an attribute is certainly a part of the concept of the attribute itself, the two are not identical. We are interested only in the conservation aspect of the concept, not in the child's *general* understanding of the attribute.

This basic distinction will be more useful if the following general model for attaining concepts about attributes is adopted. Growth in understanding of a particular attribute can be conceived of as a process of refining a concept already present. With additional refinement, the concept becomes more and more sophisticated. There has to be a starting point, a minimal concept which specifies, rightly or wrongly, when an object possesses the attribute. With respect to the attribute of length, a minimal concept would be that objects take up space, this space being related to the separation of the extremities of the object. More sophisticated concepts would recognize other properties of the expanse between extremities, e.g., that the expanse can change. The recognition of this property would constitute a refinement of the child's concept of length. Other refinements would involve the association of verbal stimuli with the concept. Such words as *big, tall, high, short* and *long* may invoke the concept. The distinction between the words *long* and *high, long* and *tall* and so on would be other refinements. A further refinement would be the recognition of the conditions under which the expanse changes, and perhaps concomitantly, the conditions under which the expanse does not change. The acquisition of these last two refinements is fundamental to conservation since, from an experimental point of view, an operational definition of conservation of length could be in terms of the child's success in specifying which actions (transformations) performed on an object will change the length, and which will leave length invariant. It is the acquisition of the last two refinements that is being tested when testing for conservation of length. Whether or not other refinements have been acquired becomes crucial only if such refinements are necessary for the valid use of the testing technique. The measure of conservation will be confounded with these other refinements if the testing technique is based partially on refinements generally acquired after the refinements being tested for.

The above formulation implies, first, that the testing technique should, as far as possible, be dependent only on refinements which occur prior to the refinements being tested for. The second implication is that in order to test any given refinement, E must communicate clearly to S the basic concept underlying the refinement.

The testing technique described in this paper was designed to fulfill these two conditions. In addition, the design was based on the assumption that children in the process of acquiring conservation of length, being somewhere between Piaget's stages of Pre-Operational and Concrete-Operational Thought, feel most comfortable and operate at their highest levels of competence when dealing with concepts embodied in *physical* objects and in *physical* action.

The Technique

It is readily apparent that the "usual" method of measuring conservation of length does not satisfy the requirements of the implications stated above. First, the method assumes that S has acquired the verbal refinement of ideas specified by the word *length*. However, a given S might associate the word length with the location of the extremities of the objects. He would say that the lengths are the same when objects are aligned in identity position (Fig. 1), but he would judge the length to have changed after the transformation was applied since the location of the extremities did indeed change. Such a S would be classified as a nonconserver. Can we be sure that he is a nonconserver, or might the conservation measure be confounded with the verbal requirements?

Second, E has no way of knowing what criteria S is using to decide whether length has changed. E can ask S why he responded as he did, but such a question assumes again that S has acquired verbal refinements related to his concept of length. The technique reported here attempts to overcome these criticisms as well as some other routine sources of error.

In general, the two criticisms can be overcome by (1) using physical apparatus and physical action to communicate to S the attribute under consideration, and (2) enabling S to give his response in a like manner. The objective with respect to (1) is to present S with a precise definition of the attribute under consideration. When the two rods are aligned in identity position (Fig. 1), the location of the endpoints and the lengths of the rods are the same. Hence the rods themselves do not clearly define the attribute under consideration. In fact, *any* attribute which has the same value for both rods in identity position could be the attribute S chooses to consider.

Since all methods for ascertaining conservation involve the process of comparison (either one object with another, or the same object with itself), and since the process of comparison is essentially the basic process

involved in measurement, the definition of the attribute could be made more precise by making the process of comparison explicit. Since the process of comparison made explicit is simply the process of measurement, the objective can be achieved by *measuring* the magnitude of length, and, as with many other cognitive phenomena, letting our measuring device define that which we are measuring.

Toward this end, one of the rods can be the measuring device, and the other the device to be measured. A crude pair of calipers can be made by attaching physical markers to the ends of the rod and affixing a small handle for ease of operation (Fig. 2). S can then be trained to use the calipers as E wishes him to.

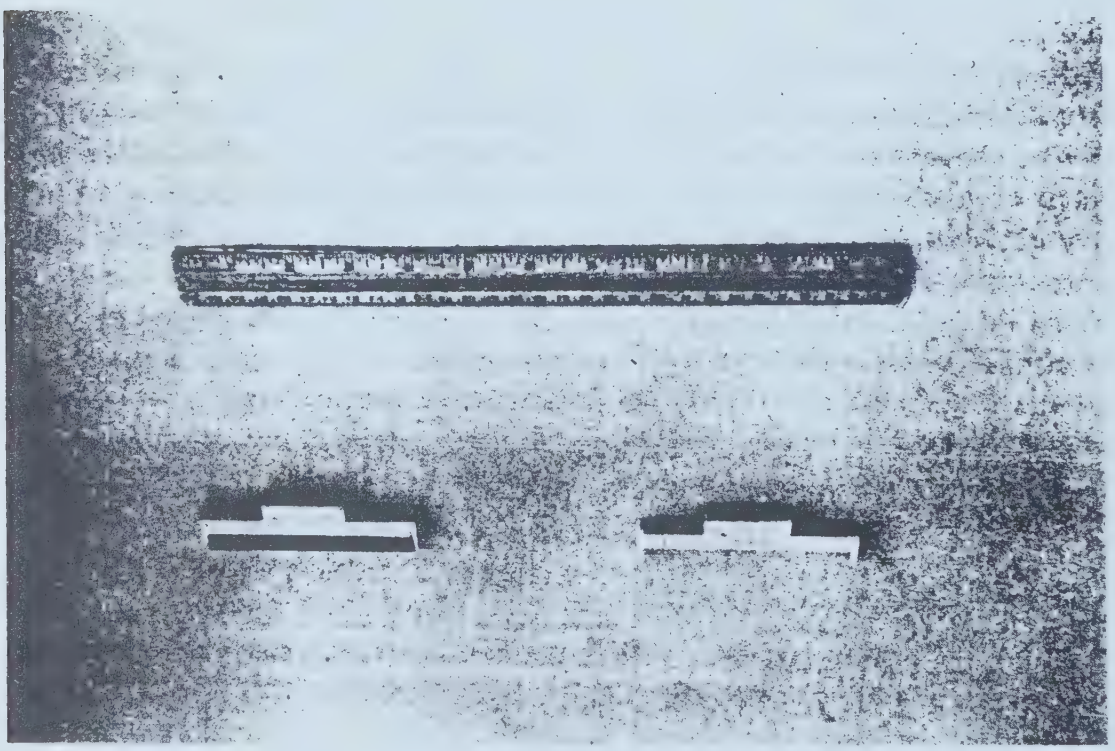


Figure 2.

Two calipers: (left) alone, (right) fitted on an object.

In such a training session, the rods could fit the calipers in only three ways. The rod could be .5 cm. too short for the calipers, just right for the calipers, or .5 cm. too long for the calipers. (Any appropriate expanse could have been chosen.) Thus, the three different fits would be defined to S. Not only would the application of the calipers give a physical definition of length, but the three different fits, differing as they do in only one respect, would also help to communicate to S the attribute under consideration. Therefore, the attribute being manipulated would also be the attribute under consideration. Zimiles[12] has pointed out that many

procedures include so much manipulation of irrelevant variables that S may not be aware of the correct criterion to use in responding. For example, the manipulation of the length of a row of checkers when the attribute under consideration is number may lead S to infer that E really wants him to respond on the basis of length.

The second major objective of the proposed technique is to enable S to use the physical apparatus (1) to facilitate his decision-making about the changes in the magnitude of the attribute, and (2) to communicate his decision to E. The three different fits previously distinguished can be used to accomplish this second objective as well as the first. A three-choice response apparatus could be constructed so that each fit is associated with one of the choices (Fig. 3). The following sequence could then be carried out. A rod and calipers are placed in front of S. S inserts the rod in the calipers (or applies the calipers to the rod) noting the fit. E then removes the calipers and applies a transformation to the rod. S *predicts* how the calipers will fit on the transformed rod by making one of the choices on the response apparatus. Embodied in his choice is his decision about the effect of the transformation on length. Once S understands what is required of him, and this can be accomplished through a training session, there need be no verbal exchange between S and E.



Figure 3.

The response apparatus with rods inserted into calipers by S.

The Experiment

In order to test the feasibility of the proposed technique, and to provide some evaluation of its effectiveness, the technique was tried on young children.

Sample

Sixty-four kindergarten and grade one children were randomly selected from a population of 196 such children attending two schools servicing the Canadian Armed Forces Barracks at Griesbach near Edmonton, Alberta. Most of the fathers were enlisted men of the Corporal or lower rank. Two of the S's had to be excluded from the study: one because he could not meet one of the criteria in the training session and the other because of illness.

Apparatus

There were three major parts to the apparatus: (1) calipers, (2) objects possessing the attribute of length, and (3) a response apparatus.

The calipers were of three magnitudes: 9 cm., 9.5 cm., and 10 cm. The objects were of three kinds—plasticine, wood, and cigarettes—and ranged in length from 1 cm. to 10 cm. The objects could easily be combined to give composite objects possessing "multiple segmented lengths". The response apparatus had 3 doors, each of which could display a candy reward when opened. As suggested earlier, the rods (objects) always fitted the calipers in one of three ways: the rod was .5 cm. too short for the calipers, the rod was just right for the calipers, or the rod was .5 cm. too long for the calipers. The association between each of these fits and the appropriate door on the apparatus was facilitated by placing a "model fit" above the door.

Procedure

Each S was tested individually by the investigator in a session which lasted from 30 to 40 minutes per S. First S was trained and then tested.

The Training Session: As can be gathered from remarks made previously, the crucial part of the technique lies in communicating clearly to S precisely what is required of him and how he is to use the apparatus to make decisions and communicate them to E. Therefore, a detailed account of the training session is given here.

The training session, which lasted from 7 to 12 minutes, had four specific objectives: (1) to train S how to use the calipers, (2) to train S to distinguish among the three kinds of fits, (3) to train S to associate each type of fit with one of the doors on the response apparatus, and (4) to train S to predict the fit of the calipers after E had given the object a transformation.

Objective 1: S was seated in front of the response apparatus (Fig. 3). S was asked if he had ever played with the material on the table and invariably S said no. "Let me show you how." E picked out a 9 cm. caliper and (4, 3, 2)* rods and placed them on the table with one of the rods on end. "We try to push these (pointing to calipers and pushing them slightly toward the rods) onto these sticks." After a slight pause, E said, "But, to make it easier we do this first. We lay the sticks down like this (upright stick is laid horizontally), and put them along here (along the front edge of the box containing the rods) nice and close (gaps are closed between sticks) in a straight line." E brings the sticks back into random formation and says, "Now you try it." Every S could arrange the rods properly. "Now see if you can push these on." S pushes the calipers on and is assisted if necessary. "Those rods fit nicely, don't they? Here, let's try these (9.5 calipers; 5, 3, 2 rods)." Several other examples were given illustrating the three different kinds of fits, thus leading to the next objective.

Objective 2: S was asked if all the sticks fit the same and invariably all S said no. S was encouraged to verbalize the differences in the three fits, but if he did not, he was not urged. If S did volunteer words to describe the fits, then E used S's terminology when referring to the various fits.

Objective 3: After distinguishing the three fits, S's attention was directed toward the three calipers lying on the response apparatus, each of the calipers with an accompanying rod placed parallel to it and about an inch from the calipers. S was directed to place each caliper onto its accompanying rod, and all S agreed that each fit was different. E asked S to tell or show him how they were different, and again E did not insist on verbal replies. S was shown the doors directly below the model fits and how they opened. The one-to-one correspondence between the doors and the fits was established by the designation by E of a fit (either by pointing or by using S's terminology) and the opening of the corresponding door by S.

E then placed calipers and rods in front of S and S applied the calipers. "Which one of these is it like?" If S designated the proper model fit, E asked, "Then which door are you going to open?" If S designated the wrong fit, he was told to look closely and try again. All incorrect designations were the result of hasty and imprecise applications of the calipers, and after trying again these S's realized that they had to apply the calipers carefully if they were to open the "correct" door and find the candy reward. Thus S was aware that close observation of the fit was necessary. S was then asked if he thought there would be candies behind the other two doors, and was invited to try them. There were no candies behind

*Denotes a 4 cm. rod, 3 cm. rod, and a 2 cm. rod.

the wrong doors. S was encouraged to verbalize as to why not, and again E did not insist.

Several other combinations of calipers and rods resulting in different fits were tried, and after S made 4 consecutive correct responses (i.e., found the candy), the final phase of the training session was begun. All except one exceptional S could meet the criterion of 4 consecutive correct responses, and all S took 7 or fewer trials ($p = 4/81$, at a maximum). The exceptional S, who had a very low IQ, was excluded from the study.

Objective 4: At this stage, all S agreed that it was easy to find the candy, as indeed it was. "Well since it's so easy, I'm going to change the game so that it's not quite so easy, and this time you can keep the candy when you find it. I'll show you what we're going to do; watch closely so that you'll be able to find the candy. We're going to do three things. First you put the calipers on the sticks just as before (S does so). Next I'm going to do something to the sticks (E removes the calipers and transforms the sticks). Now *if* you put these (calipers) on again in the way you always do, would they be like this, or this, or this (pointing to model fits)?" Some S pointed at a model fit, but some started to put the calipers on again. For those who pointed at a fit E said, "Where's the candy going to be then? Find it." If S found the candy he was told, "Yes, that's right. That's what it would look like." If S failed to find the candy (i.e., designated the wrong fit) he was told, "No, I guess it wouldn't be that one. It would be like this one or that one." (Parenthetically, it should be noted that up to this point the general objective has been to specify to S that length is the attribute under consideration. It now remains to specify to S that the refinement of conservation is the crucial variable.) For those S who started to put the calipers on the rods, E interrupted and said, "That would make it too easy if I let you do that. I want you to tell me what it would look like without having to put these on. Would it be like this, or this, or this?"

All S were required to do additional examples and all S appeared to grasp what was required of them. In the additional examples E inserted the sentence, "You saw what I did to the sticks. What's it going to look like now, this, or this, or this?" In an attempt to prevent the child from getting the idea that he merely had to choose the original fit again, two of the transformations used changed the length of the rods. The test itself was then begun.

The Test

The 24 test items were given in exactly the same manner as the last examples in the training session. Sixteen of the transformations left length invariant, while eight changed the magnitude of length. The transformations consisted of such actions as (1) rotating the rod(s),

(2) translating the rod(s), (3) adding or subtracting a portion of the rod(s), (4) rolling the plasticine rod, thus elongating it, and (5) cutting a portion of the filter off a cigarette (or adding it on).

Administration of the test: An attempt was made to minimize the confounding influence of several variables:

1. Guessing. In a multiple choice context, guessing plays a prominent role. To counteract its effect, each of the three doors led to a candy reward an equal number of times (8, 8, 8); twenty-four items were used instead of the usual three or four; and S was told that if he was careful and watched what E did to the rods, he could find the candy *every time*.

2. Perceptual estimation. To counteract perceptual estimation the calipers were at no time left in a position parallel to the rods. After removing the calipers, E usually held them partially concealed in his hand, or laid them perpendicular to the rods when possible.

3. Disinterested performance. The use of candy rewards, the novelty of the apparatus and of the technique, along with the active involvement of S throughout the testing session helped to overcome lack of interest.

4. Forgetting. It was crucial that S remember how the rod fit before the transformation was applied. To minimize the role of forgetting, any time that S appeared to forget the initial fit E said, "Did you forget how it fit before? It looked like this one. You saw what I did to the sticks. Which one do you think it's going to be now? Find the candy." Moreover, S was informed that if he ever forgot, he should ask E.

Results

In order to assess the feasibility and effectiveness of the technique it was assumed that if the technique did in fact remove several confounding influences, then the removal of such confounding influences should be reflected in the increased success that younger children would experience with the technique as compared with the success they would have with the usual method. In discussing the age of acquisition of concrete-operational thought processes, Braine[13,14] and Smedslund[15,16] reported their results as threshold ages determined by finding the age interval for which 50% of the subjects in the interval possess the thought process. On this basis they estimated, for Piaget's[17] results, the threshold age for the acquisition of conservation of length to be between 7 and 8 years of age. Murray[18] reported similar threshold ages. Thus, on the basis of the rationale presented here, the effectiveness of the proposed technique can be determined by testing the hypothesis that the technique will result in a significantly lower threshold age for the acquisition of conservation of length.

Before any results concerning the threshold age can be reported, criteria to determine when a given S can be considered a conserver must

be specified. In most tests of conservation, the decision-making process which confronts S can be conceived of as essentially a process of classification. S has to classify the transformations applied to the objects into two mutually exclusive categories: those transformations which leave length invariant and those transformations which change the magnitude of length. In this kind of situation, guessing would give 50-50 results. For the technique used in this study, the involvement of three different fits made possible a finer classification of transformations: those which decreased length, those which left length invariant, and those which increased length. A conserver was one who could classify the transformations correctly in these three categories.

Assuming that all important confounding influences were controlled, and since a three-way response apparatus was used, the probability of a correct response from a nonconserver is no more than one in three. For the purposes of reporting results, the 16 test items which left length invariant were divided into two sets of 8, one involving rotations and the other translations. Each of these sets of 8 was combined with the 8 items which did not leave length invariant. Thus, there were two overlapping subtests of 16 items each. Since half of the items in each subtest changed the magnitude of length, the probability was minimized that any S would maintain the strategy of responding on the basis that the fit after the transformation would be the same as the fit before the transformation.

With $\theta = 1/3$ and $N = 16$, a score of 11 correct responses would be significantly better than chance at the .01 level using the binomial distribution.

Table 1 shows, for each subtest, the number and percentage who reached the criterion in each age interval. Data presented in the table indicate that the threshold age for conservation of length, using the proposed technique, is somewhere between 5-4 and 6-2. These ages are about 2 years lower than those obtained using the usual technique. The hypothesis concerning the effectiveness of the proposed technique is therefore accepted.

Discussion and Summary

A modified technique for assessing the child's acquisition of the conservation of length was presented and evaluated. An attempt was made to control for the following confounding influences: guessing, perceptual estimation, misunderstanding of instructions, misconception of the response criteria, failure to perceive the initial comparison of lengths (perception of the initial "fit"), forgetting the initial comparison, lack of a nonverbal way of expressing decisions, and disinterested performance.*

Perhaps the strongest modification in the proposed technique is the use of physical apparatus to make the technique highly concrete and

*Based on Smedslund's analysis (1963).

more precise, thus overcoming the vagueness and imprecision associated with children's language. However, a price must be paid for introducing such precision, the price being embodied in the logical relations built into the use of the apparatus. The crucial question is whether the dependence on a minimal logical competence is more confounding than the dependence on a minimal verbal competence. Since the experiment resulted in threshold ages about 2 years lower than the usual ages, it is concluded that when the required logic is embodied in concrete operations with physical apparatus, the young child is better able to reveal his threshold acquisition of cognitive processes.

TABLE I

ACQUISITION OF THE CONSERVATION OF LENGTH WITH RESPECT TO AGE

Age Group	No. of Ss	Subtests Serving as Criteria			
		Rotations and Non-Inv		Translations and Non-Inv	
		N	%	N	%
7-2 to 8-0	15	14	93	15	100
6-3 to 7-1	22	19	86	17	77
5-4 to 6-2	25	19	76	16	64

Note: 1. Ss were classified as conservers if the number of correct responses they gave on a particular criterion was significant at the .01 level.
2. "N" designates the number of Ss with conservation.
3. "%" designates the percentage of Ss with conservation.

REFERENCES

[1] A. T. Brace, *The Pre-School Child's Concept of Number,* Unpublished Master's Thesis, The University of Alberta, 1963.

[2] P. C. Dodwell, "Children's Understanding of Number and Related Concepts", *Canadian Journal of Psychology,* 1960, 14, pp. 191-205.

[3] K. E. Feigenbaum, "Task Complexity and IQ as Variables in Piaget's Problem of Conservation", *Child Development,* 1966, 37, pp. 977-983.

[4] I. C. Uzgiris, "'Situational Generality of Conservation", *Child Development,* 1964, 35, pp. 831-841.

[5] E. Halpern, "The Effects of Incompatability Between Perception and Logic in Piaget's Stage of Concrete Operations", *Child Development,* 1965, 36, pp. 491-497.

[6] M. D. S. Braine, "The Ontogeny of Certain Logical Operations: Piaget's Formulation Examined by Nonverbal Methods", *Psychological Monographs,* LXXV, 1959, No. 5.

[7] "Piaget on Reasoning: A Methodological Critique and Alternative Proposals", in *Thought in the Young Child,* William Kessen and Clementina Kuhlman (Eds.), Monograph of the Society for Research in Child Development, 1962, 27, pp. 41-61.

[8] "Development of a Grasp of Transitivity of Length: A Reply to Smedslund", *Child Development,* 1964, 35, pp. 799-810.

[9] J. Smedslund, "Development of Concrete Transitivity of Length in Children", *Child Development,* 1963, 34, pp. 389-405.

[10] "The Development of Transitivity of Length: A Comment on Braine's Reply", *Child Development,* 1965, 36 pp. 577-580.

[11] G. E. Gruen, "Note on Conservation: Methodological and Definitional Considerations", *Child Development,* 1966, 37 pp. 977-983.

[12] H. Zimiles, "A Note on Piaget's Concept of Conservation", *Child Development,* 1963, 34, pp. 691-695.

[13] Braine, *op. cit.,* 1962.

[14] Braine, *op. cit.,* 1964.

[15] Smedslund, *op. cit.,* 1963.

[16] Smedslund, *op. cit.,* 1965.

[17] J. Piaget, *The Child's Conception of Geometry,* (New York: Harper & Row, Publishers, 1960).

[18] F. B. Murray, "Conservation of Illusion-Distorted Lengths and Areas by Primary School Children", *Journal of Educational Psychology,* 1965, 56, pp. 62-66.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 425-425*

Erwin Miklos
Editor, December 1969-June 1972

Individualizing Learning

A significant development in educational research in Alberta occurred in the late 1960s with the establishment of the Human Resources Research Council. Although the mandate of the Council included all types of social science research, the major initial thrust was the research and development program in education. A specific aspect of the program focused on exploring means of individualizing learning. In the Individually Prescribed Instruction (IPI) Project, materials for mathematics developed at the University of Pittsburgh were field tested in three schools. The venture into research and development brought with it the need to design ways of evaluating such activities.

In his informative article "Evaluation of the IPI Project," Maguire describes the comprehensive approach taken to evaluating the Project. At the time, the challenges of conducting evaluation research were well documented in the literature. But examples of how sound evaluations might be conducted under less than ideal conditions were lacking. The article by Maguire helped to fill the gap and continues to stand as an example of how to conduct evaluation research. The approach taken is noteworthy for the breadth of perspective that encompasses both process and outcomes, for the comprehensive methodology that includes both quantitative and qualitative strategies, for the attention given to collecting data from wide-ranging sources, for the use of varied data collection techniques, for the thorough manner in which the results are presented, and for the balanced and thoughtful interpretation of results. The evaluation must have been helpful to all stakeholders in the project as well as to other researchers.

No doubt the article would have been even more influential on the local scene if the research and development activities such as the IPI Project had not ended abruptly because the Human Resources Research Council failed to survive a change of provincial governments in the early 1970s.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 426-444*

THOMAS O. MAGUIRE
The University of Alberta
and
Evaluator, IPI Project
Human Resources Research Council

Evaluation of The IPI Project

The evaluation of the IPI mathematics project in three Alberta schools is described. Data were collected on student achievement, student attitudes, teacher attitudes and parents' opinions. In a comparative study of student achievement, the results tended to favor the set of Control schools. In areas of student and teacher attitudes the reverse was true. Parental opinions were found to be generally favorable to the program. Project administrator reports are described and implementation problems are discussed. The present paper is a much-shortened form of a report presented to the Alberta Human Resources Research Council.

In the spring of 1969, three schools were selected as sites for the establishment of demonstration/testing centres for IPI Mathematics. Two urban schools were selected as representative of two broad socio-economic classes. One of the schools served a predominantly lower middle class neighbourhood while the other urban school was located in an upper middle class area. A third school was selected to represent the rural situation. The three school systems in whose jurisdiction the demonstration schools fell were asked to nominate a second school which most nearly matched the demonstration school in terms of facilities and area served. In the report that follows, these schools will be referred to as Control schools, although it is important to realize that from an experimental point of view, they could not be considered as equivalent to the IPI schools in any more than a general way.

In September of 1969, IPI mathematics was installed in the three IPI schools. As the project developed the major emphasis was placed on the third objective. In line with this, the evaluation activities were also concentrated on investigating the applicability of the IPI Mathematics system to Alberta schools. To this end, during the first year the evaluators undertook to examine IPI mathematics from two points of view, process and outcome.

Reprinted from *The Alberta Journal of Educational Research*, 17(4), 255-273, 1971.

Two broad procedures were used in the evaluation. Comparisons were made between the IPI and Control schools on many of the variables. In addition, and perhaps more importantly, observations, opinions and judgements were elicited from students, teachers, parents and administrators in the IPI schools.

Achievement data, and some attitude data were collected in October of 1969, and again in May of 1970, in an effort to detect changes that might have occurred during the first year of the project. Interviews with teachers and parents were made during February of 1970, after they had a chance to get by the initial disruption of installation, and had experienced enough of the program to form valid opinions. All of the results presented in this report refer to the first year of the project.

Process Evaluation

Two kinds of processes were involved in the IPI Project. The first process consisted of the installation of IPI in the schools. The data concerning this process were the observations and opinions of two of the administrators who had responsibility for running the program. A qualitatively different kind of process was investigated by looking at the activities in the classroom themselves. A comparative study of the teacher-student interactions was made to see if any differences in classroom processes were detectable.

Evaluation of Program Implementation

Pacey (1971a) and Holmes (1971), two of the administrators involved in the implementation of IPI, maintained sets of notes describing the problems and successes that were encountered in the application of IPI during the first year. From their reports, several interesting results were noted.

During the first month of operation, all of the children had to be placed on the IPI continuum. This required the testing of all of the children to determine the level at which they should enter the program. The procedure proved to be tedious and often exhausting, but since complete placement testing is only required when the program is first introduced to the school, the task was a bearable one. In succeeding years, September placements for most students can be made on the basis of their June locations on the continuum.

Also, in the first month, the schools needed extra help in carrying out the many tasks of recording, prescribing, testing, marking, and distributing materials. Because the teachers were not yet intimately familiar with the materials, more than the usual help was needed. The school administrators spent a considerable amount of time in the classroom working with the teacher and students in an effort to alleviate the problem. This not only provided the extra manpower necessary to overcome the initial starting pains, but had the side effect of promoting a spirit of unity among the staff.

Because of the magnitude of the individualization task, the first month was particularly inefficient for the students as well. Some students spent as much as three quarters of their time standing in line, waiting to have materials corrected or assigned. Two solutions to this problem were used. The children themselves were given more responsibility in marking their own work, and obtaining materials from the storage rooms. Secondly, since the city schools

obviously needed more teacher assistance for the clerical chores, parent volunteers were obtained from the communities. Typically two volunteers were used in each math class.

Among the other initial difficulties was the problem of the placement tests. Children who had grown accustomed to the traditional purposes of testing, often had difficulty accepting "failure" in placement testing. However, after a month's experience with the program, their understanding of the role of the placement test allowed them to see it in its proper perspective.

The administrators of the IPI program in each school were trained at a demonstration school in the United States. At that school, the IPI guidelines were rigidly adhered to. Within the Alberta situation, no attempt was made to follow the guidelines exactly. The teachers of each school tried to fit the program into the contexts of their own situation. In grade one for example, one school used the IPI materials from the outset, but used additional mathematics readiness materials at the same time. At another school the children in grade one were phased into the program in three groups as they became ready. In general the teachers tried to follow the IPI program, but felt free to deviate from the enunciated procedures whenever their professional judgement dictated it.

The effects on the children, as observed by the administrators was important. Almost all of the children were excited by the program. Undoubtedly some of this could be attributed to the novelty of the program, but that is not necessarily a bad thing. There were no new discipline problems caused by the increase in freedom, and some of the old problems (a few of the so-called lazy children) seemed to disappear. Some of the children were bewildered and frustrated by the complex system, but the number was not nearly as many as the staff had expected. The vast majority of the students adjusted quickly and enjoyed their work so much that they were reluctant to be taken out of their booklet work to work on some enriching activities in seminars. There seemed to be a lingering competitiveness that made some of the children push to keep ahead of their friends.

For the teachers, the work at the beginning was hard. Later, as they became more accustomed to the system, they began to sense a different kind of problem that was caused by the IPI system. In traditional group teaching situations, individual learning problems are often overlooked in the attempt to achieve some overall mastery by the majority of students. In IPI, the problem does not go away as the student moves from one topic to another, because he cannot move until he has achieved the goal. As a result, teachers were forced to deal with individual learning problems and were astounded by the complexity of the learning process. Despite the difficulties, the teachers were generally positive in their feelings toward the program, although they were critical of many of the specifics of the curriculum.

Five general problems were detected by the administrators.

1. The curriculum materials that were used to teach the money units, all used American coinage; these units were changed to fit the Canadian scene.

2. The project was plagued with problems in shipping, both in terms of late shipments, and improperly filled orders.

3. There were many printing errors in the materials and answer keys.

4. the units designed for the first grade students, required a greater ability to read than most students possessed.

5. Uncertainty in the organization of the funding agency was disruptive of morale in the schools.

In the minds of the participant-observer administrators, these five problems were more than compensated for by the enthusiasm of the learners and teachers.

Classroom Interaction Evaluation

Some critics of IPI suggest that because of the individualized nature of the program, the children loose valuable opportunities for developing social skills. In an effort to determine what differences in classroom interactions existed between the programs, Pacey (1971b) carried out a comparison of the verbal behavior patterns in the experimental and control schools using the Flanders method.

All teachers in grades three to six in all six schools were observed for twenty minutes each. In all there were 16 hours of observation time. The observations were made in March and April of 1970.

The analysis of the interaction matrices indicated several differences between the two groups. The results were summarized by Pacey as follows:

1. More information was provided directly by the teacher in IPI. Closer examination revealed that this tendency was general throughout the matrix. That is, IPI teachers tended to precede and follow student talk with their own thoughts and opinions more than did teachers in the control group. There was no other significant build-up elsewhere, for example, use of extended lecture was about the same in both groups. This indicated that rather than taking time to build on student ideas, IPI teachers tended to explain answers quickly and move on, which may suggest that the IPI system is rushing teachers.

2. IPI teachers made less use of questioning. The pattern for this is consistent with number one. These two differences would seem to be significant because the IPI system claims that teachers should avoid telling students answers, but rather should encourage them to discover answers for themselves, by leading them there via careful questioning.

3. There was substantial increase in student talk in the IPI group, especially student-initiated talk. This is certainly, at least in part, due to the system, in which students come to the teacher for assistance or guidance, usually at their own initiative. But careful analysis of teacher reaction to student talk indicated that the teachers in the IPI group seemed to encourage it.

4. Related to increased student talk was the dramatic increase in extended student talk. When a student began talking to the teacher, he was allowed to talk for sustained periods of time more frequently in the IPI group than in the Control group.

5. A significant increase was evident in the use of motivating and encouraging teacher behavior in the IPI group. Although use of these behaviors was not very extensive, it was higher than in the Control group. This, of course, may be due to differences between the schools, but it might also be

attributable in part to the IPI system, which allows one-to-one interaction on a far greater scale than normally.

6. There was a corresponding decrease in the use of controlling teacher behaviors in the IPI group. That is, IPI teachers tend to give directions and criticism less often than did the Control teachers. The IPI system would seem to account for the decrease in giving directions, since the materials do much of that. The decrease in criticism might be partly attributable to the system, which is success-oriented and gives students a fair amount of freedom.

7. There was less use of extended indirect teacher behavior in the IPI schools than in the Control schools. This may suggest once again that teachers do not have enough time to adequately develop and encourage student ideas and feelings.

Because IPI had only been in the schools for six or seven months at the time of the observations, the study was viewed as exploratory in nature.

Outcome Evaluation

The outcomes of the program were examined in two ways, comparatively, and judgmentally. Two general kinds of effects were investigated, achievement, and attitude. The attitude effects were investigated with students, teachers, and a sample of parents.

Comparative Study of Student Achievement

In an effort to evaluate student achievement in mathematics in a comparative way, two subjects of the Canadian Tests of Basic Skills (CTBS) were administered to the students in grades 3, 4, 5 and 6 in the IPI and Control schools. The CTBS is a version of the Iowa Tests of Basic Skills adapted for use in Canadian schools. The two subtests used in the evaluation were Arithmetic Concepts and Problem Solving. These tests were designed to reflect the skills and objectives pursued in most existing textbooks. The Arithmetic Skills test contains items measuring concepts involving currency, decimals, equations, fractions, geometry, measurement, numerals and number systems, percents, ratios, and whole numbers. The Problem Solving test includes problems that involve the kinds of operations and concepts that generally appear in the first half of textbooks for a particular grade.

The tests were administered in September and May of the first year of the program. At each grade level an analysis of covariance was carried out on both the Arithmetic Concepts and Problem Solving tests to see if there were significant differences in the May means after adjustments were made for differences that existed in September.

The results of the analysis for both tests showed no significant differences at grades 4 and 5. For grades 3 and 6, the adjusted mean for the Control group was significantly higher than that of the IPI group.

In all cases, the differences in adjusted averages, did not exceed 2.25 points. The individual school averages for the May testing were compared to the norms for school averages provided by the publishers of the test. These norms indicate the percentile points for grade averages when the tests are administered after April 1. For the six schools in the study, and four grade levels in the study, only two grades of the 24 were below the 50 percentile

on Arithmetic concepts, and only five grades were below the 50 percentile on Problem Solving. This record indicates that all of the schools seem to be doing a remarkable job (in relation to the schools used in the norming procedure) in preparing their students in areas measured by the CTBS. Thus, even though differences that exist favor the Control group, this should not be taken to mean that the results for the IPI group are poor.

In an effort to determine whether or not there were any differential effects of the program for different ability levels, the data were reanalysed with the students categorized into ability levels. The Lorge Thorndike Verbal and Non-Verbal Intelligence. Scales were used separately to categorize the students into three broad intellectual groups. Because the test was administered at the same times as the CTBS, it was possible to place students into the three groups according to their September and May scores. Only data for students whose I.Q. scores fell within the same category on both administrations of the test were used in subsequent analyses. The categories were defined as I.Q. scores of 95 and below, 96-115, and above 115. Once the students had been classified by ability, students from each category were selected at random until there were equal numbers of students in each category defined by the IPI-Control dimension and the high, medium, low ability dimension.

In all analyses, the September score on the CTBS scale was used as a covariate in an attempt to partial out differences that existed among the groups on the September testing. In one way, the analysis is really an analysis of gains made over the year. In the analysis, it was expected that the higher ability levels would gain more than the lower ability levels so that one would expect a significant difference among the ability levels. Differences between the IPI and Control groups might also exist as had been noted in the previous analysis, although some inconsistencies with the previous analysis might exist because of the smaller number of students used in the present analysis. The critical test, and the purpose of the present analysis was the test for the interaction between program and ability level. Presence of a significant interaction might indicate the effects of the IPI and Control programs on one ability level were different from the effects at another level.

The results of the analysis showed significant differences among ability groups, as expected, and between IPI and Control groups at grades 3 and 6 as shown previously. There were no significant interactions between ability groups and program. Since no interactions were significant, there is no evidence to suggest that either program is more effective at one level than at another.

Of course, the success or failure of a program cannot rest on a single or even two measures of educational achievement. Because the IPI program is generally conceded to be an innovation in method and not content, the CTBS Scale used seemed appropriate measures for determining gross differences between the IPI and Control schools. However, future evaluation should include finer instruments in order to be able to document specific differences between the groups.

Where differences did occur, (at grades 3 and 6), the Control group showed superiority. This may be due to a number of factors, each of which

must be considered for future evaluation and decisions. While the schools' curricula did not differ greatly in content, the sequence of material may have been able to account for the differences that occurred. The IPI sequence may have put them at a disadvantage in grades 3 and 6, or, on the other hand, the Control schools may have been at a disadvantage in grades 4 and 5. Another factor that may have accounted for the results is that the students at the different grade levels were different for the two programs. The analysis of covariance in the study was used in an attempt to overcome differences in CTBS scores that existed prior to the implementation of the program, it cannot be expected to make the groups "equal" however. Differences in teachers and classroom atmosphere could also have contributed to the results. Finally, it is also possible that the IPI program itself is not as strong at grades 3 and 6 as it is at grades 4 and 5.

Measurement of Student Attitudes

As one method of collecting information about changes in student attitudes over the first year of the program, a Student Questionnaire was administered to the students in grades 3 to 6 of the IPI and Control schools in September and May, (Maguire, 1971).

In general, the results of the questionnaire indicated that the students in the IPI schools at all grades liked mathematics better than their counterparts in the Control schools. A second finding was that more students in the IPI schools found mathematics easy, than in the Control schools. Both of these findings occurred at September and May which leads one to believe that there are additional factors operating beyond the IPI program itself. However, since the September administration was made after the children had been in school for almost a month, the month's experience may have caused the difference.

In addition to collecting attitudes toward arithmetic activities an attempt was made to evaluate some generalized student attitudes toward their environment. In May, a "pupil opinionnaire" developed by the staff at Research for Better Schools Inc. for use in their evaluation of student attitudes toward IPI was administered. This opinionnaire used a semantic differential format. Students were asked to rate eight concepts on 9 five point scales. The eight concepts were:

My School Is

Being Praised for Doing Something Well is

My Classmates Are

Working by Myself in School is

My Math Class is

Being Scolded for No Reason is

I am

Taking a Math Test is

The nine scales used were:

smart—stupid
happy—sad
easy—hard

fair—unfair
bad—good
slow—fast
nice—awful
useful—useless
boring—interesting

The instrument was administered to all students in grades 3, 4, 5 and 6. The responses were made anonymously. Because the instrument was being used on a speculative basis, the students were not required to indicate their grade level. As a result there were 497 responses in the IPI group and 542 responses in the Control group.

The differences between the average ratings of the IPI and Control groups were tested for each concept on each scale. In all there were 72 tests carried out.

For the concepts "Being praised for Doing Something Nice", "Being Scolded for no Reason", and "I am", there were no significant differences on any of the scales. For the "My Classmates are" concept, the IPI group rated their classmates closer to the smart, fast, and good ends of those scales than the Control group. All other scales were not significant.

The "My School" concept showed significant differences on all scales except hard-easy. The "My Math Class' concept showed significant differences on all scales except useful-useless, where the average rating for both groups was next to the useful end of the scale. "Working by Myself in School" was significant on all scales except hard-easy and fast-slow. On all of the scales on which significance occurred, the IPI group means were closer to the smart, fair, interesting, nice, good, happy, easy, fast, useful ends of the scales than were the Control means. In all of these cases too the Control group was closer to the above ends of the scales than they were to the other end.

For "Taking a Math Test", the IPI group rated it closer to the nice, interesting and happy ends of those scales. There were no significant differences on the other scales.

The interesting findings is that despite the extremely powerful test used (the degrees of freedom were over 1,000) not all of the tests produced significant results. In tests that involved general concepts such as "Being Praised . . . and "Being Scolded . . ." no significant results occurred. These results suggest that the scale may have some validity and that the IPI children are not trying to produce results that merely please. The self concept scale, "I am", also showed no significant scales. The "Classmates" concept only showed three significant scales. In general, the concepts involving school experiences that were different for the two groups did produce significant results with the IPI group exhibiting some positive responses. Of course, one must note that the Control group responses were not negative. In most cases they were only slightly less positive.

As a result of the two questionnaires, it seemed fair to conclude that the IPI program as installed in the three schools of the study tended to produce a greater enthusiasm for arithmetic among the students than the Control program did for its students. This enthusiasm does not appear to be blind. On questions that do not reflect program characteristics, there appear to be few differences between the groups.

It is difficult to separate attitudes towards mathematics from attitudes towards the mathematics environment. We cannot conclude that IPI caused student enthusiasm. It may have been skillful teachers who produced this result and they may have done so using any innovative program. In future evaluation activities it will be useful to interview students so that the questions can be explored.

Comparative Study of Teacher Attitudes

In order to tap the general feelings that teachers in the IPI and Control schools have toward a number of Educational topics, an attitude instrument was constructed using the semantic differential format. Two kinds of topics were covered. The first set of topics: Innovation, Mathematics Education, and Individualized Instruction, were chosen because of their direct relevant to the project. The second set: Students, Teachers, Parents, and School System were included in an effort to obtain some measure of the general level of morale among the teachers.

The teachers in both groups were asked to rate the topics on the 29 adjective scales shown below. The instrument was administered in October and May.

Exciting—Dull	Systematic—Chaotic
Temporary—Lasting	Extravagant—Necessary
Authoritarian—Democratic	Worthwhile—Futile
Society oriented—Individual oriented	Ineffective—Effective
Traditional—Modern	Reflective—Impulsive
Difficult—Easy	Isolated—Integrative
Practical—Theoretical	Stationary—Moving
Enjoyable—Boring	Peripheral—Basic
Important—Useless	Inexpensive—Expensive
Concrete—Abstract	Frustrating—Satisfying
Good—Bad	Positive—Negative
Weak—Strong	Inadequate—Adequate
Precise—Vague	Not time consuming—Time consuming
Divergent—Convergent	Active—Passive

Two way analyses of variance were carried out on the judgements of each concept on each scale. The independent variables used were IPI versus Control group and October versus May testing. In all, 203 such analyses were carried out. Since there may be some very high correlations among scales, it is very likely that many of the results that are judged to be significant could have occurred by chance (even if the scales were independent, one would expect 5% of them to be significant.) Nevertheless, for the sake of simplicity of presentation and interpretation, the two way analyses were carried out. Most of the teachers responded anonymously, so that it was not possible to analyse the results properly, that is by taking into consideration the correlation between pre and post tests.

In considering the results, it is useful to keep in mind the rationale for carrying out the test. If participation in the IPI program (or factors concurrent with it such as Hawthorne effect) produced changes in teachers' attitudes that were different from changes that occurred in the Control group then a significant interaction should result. If there were a general change in attitude for both groups, then a pre-post test (Time) effect would be significant. If

the groups differed in attitude prior to introduction to the program, and this difference persisted throughout the program, then a significant Program effect would be noted.

In almost all cases, it was clear that even though one group rated "Innovation" for example, as being duller than the other group rated it, the average values for both groups fell on the exciting side of the scale. Thus such a result cannot be interpreted as one group's thinking that Innovation is dull and the other thinking that it is exciting, rather it is an indication that one group sees innovation as being more exciting than the other.

The results showed that only two interactions occurred. This seems to indicate that any differences between the groups may have existed prior to the first testing occasion.

Part of the result that the teachers' attitudes were different prior to the first testing occasion may have arisen from the intensive training program that was conducted. At a week-long course in IPI methodology that was held in Baff in August 1969 much enthusiasm was generated among the teachers for the program. The attitude scale was administered in late September and early October.

A second confounding factor was the degree of attention that was paid by the Control group to the scale. The Control group received no benefit from their participation in the program. In general the teacher attitude scales were administered after the Control group had been subjected to a batch of other tests in their classrooms and while the cooperation was good, one could not help but feel that the attitude scale was seen as an unnecessary imposition on the teachers' time. No indication of any attitude toward the questionnaire was received from the IPI group.

If one looks at the concepts individually, a picture of the teacher's attitudes begins to form. For the concept Individualization, the IPI teachers tended to see it as more exciting, lasting, democratic, individual oriented, modern, practical, enjoyable, important, good, strong, precise, integrative, moving, satisfying, worthwhile, effective, reflective, positive, adequate and active than the Control group. Again the responses indicate that the IPI teachers are generally enthusiastic about the program. It is interesting to note there were no differences on the time scale. This is one of the situations where both groups saw individualization as very time consuming. Similarly on the difficult-easy scale, both groups placed Individualization at the difficult end.

The Parents' concept showed significant differences on 17 scales, with the IPI teachers seeing parents as more exciting, lasting, democratic, easy, practical, theoretical, enjoyable, important, good, integrative, inexpensive, satisfying, worthwhile, effective, positive, adequate, not time consuming, and active than the Control group. The responses to this concept illustrate one of the problems with this technique. Often seemingly inappropriate scales such as temporary-lasting show differences among groups. With interviews, the reasons for the responses can be explored; with questionnaires, the reasons are left to the imagination of the interpreter. Two kinds of interaction with the parents were different for the two groups. Parent interviews were used in place of report cards for the IPI group, and parent volunteers were used to aid the teachers in the classroom.

For some of the scales it may have been this increased contact with parents that produced the results. In particular the responses to the scales such as enjoyable, satisfying and worthwhile might reflect this. On scales such as lasting and inexpensive, the IPI group may be referring to the effect that parents were having on the IPI program as a result of their volunteer activites.

Sixteen scales were significant for the concept Students. The IPI teachers rated this concept as more exciting, lasting, democratic, individual oriented, modern, practical, enjoyable, important, strong, precise, moving, satifying, systematic, worthwhile, positive and adequate. Again, it is difficult to separate generalized reactions to students from reactions to students within the context of IPI. It seems safest to suggest that all of the teachers seem to have a high regard for students with IPI teachers giving slightly higher ratings on some dimensions. This may have resulted from more contacts with students on a one-to-one basis since the IPI system requires teacher-pupil conferences. Whether or not the actual time spent with individual students is greater in the IPI system has yet to be established.

IPI teachers rated Teaching as more exciting, lasting, democratic, individual oriented, modern, enjoyable, important, integrative, moving, effective, positive, adequate and active than did their Control counterparts. Both groups saw teaching as being practical, good, strong, complex, expensive, satisfying, systematic, necessary, worthwhile, and time consuming. It would seem that the teachers in both groups must be described as dedicated to their vocation. Large differences existed on two scales: democratic and individual oriented. These scales may reflect attitudes that are program specific.

Mathematics Education produced only nine significant scales. IPI teachers rated the concept as more exciting, democratic, individual oriented, integrative, moving, expensive, effective, and active. Many of these scales seem fairly closely related to the program. The attitudes toward Mathematics Education in a broad sense seem fairly similar.

The School System, the concept furthest from the IPI program, showed significance on six scales. The IPI teachers rated it more democratic, important, necessary, worthwhile, reflective and adequate. Perhaps, six scales represent a sort of chance level. None of these scales seems particularly important. In general, teachers' reactions to the School System seem to vary from neutral to fairly positive. One would not expect many differences here as the School System is quite remote from individual teachers.

Given the conditions under which the instrument was administered; that is after the program had been in operation for a month, one can see that it would be difficult to show the significant interactions suggested earlier. It does appear that the enthusiasm toward the program has not flagged over the year between administrations. In looking in general terms over the concepts and scales it seems fair to conclude that the IPI teachers have responded enthusiastically to the research that has been conducted in their schools. Much of this effect is likely due to either the novelty of the situation, or the special treatment given

to the schools, however, collection of similar data in subsequent years when the novelty effect has gone will indicate long term differences that exist in attitudes.

Teacher Interviews

Conners (1971) interviewed the 33 teachers who had worked with IPI for eight months in an effort to determine their opinions about the program. Six question areas were explored.

Question One: What perceptions do teachers have of IPI mathematics as an instructional package?

The study revealed that most teachers were extremely satisfied with IPI mathematics as an instructional package. The objectives of the program, the materials used, as well as the structure and flexibility of the program were considered as good, if not superior to most programs. However, more definite advantage were claimed for IPI over most other programs. In the opinion of the majority of teachers, IPI mathematics provides better for the "above average" and "average" pupils than most mathematics programs because it does not inhibit the pupils' progress. IPI also, through its variety of instructional approaches, appears to provide well for the different learning styles of pupils and allows pupils to work at their own pace. It must also be noted that after contact with IPI, all teachers in this study indicated their desire to individualize instruction in most other subject areas.

While IPI appears to have many "strengths," there appear to be some "weaknesses" in the program. The data suggest that IPI does not free the teacher from teaching the routine, basic skills of mathematics. It appears that efforts on the part of the teacher have now been transferred from a large group situation to an individual basis. The data also suggest that IPI, because of the structure and the pressures of the system, does not provide "well" for the student who wishes to pursue some special interests in mathematics.

Question Two: What perceptions do teachers have of the role of the teacher in the IPI system?

The data suggested that in the IPI system the teacher's role is changed from the traditional concept. The teacher lectures far less, works harder and has more demands made upon him in knowing the subject matter and curriculum. However, it appears that there is greater satisfaction working in an IPI classroom when compared to a conventional classroom because the individual contact with students allows the teacher more opportunity to assist the individual pupil than was previously possible in a conventional classroom. This satisfaction is enhanced because of the immediate feedback concerning the student's progress. The data also indicated that teachers in the IPI situation find their role to be more challenging, modern and professional than in a conventional classroom. It appears that IPI presents an opportunity for the teacher to work effectively with each individual in the classroom. The teacher is given an opportunity to become more "professional" as she becomes a "diagnostician" of learning, a decision maker and a guide for the pupil rather

than a "fund of knowledge." It appears therefore, that IPI creates greater opportunities for the teacher to use all her science and art as a teacher, in order to assist the individual pupil.

The data also suggest that contrary to some opinions, the IPI teachers do not find working in the IPI situation to be a "dehumanizing" experience for themselves or the pupils. In fact, it is suggested that IPI assists teaching and learning to become a more "humanizing" experience for both teachers and pupils because the teachers come to know and understand the students better as individuals.

Question Three: What perceptions do teachers have of pupil reactions to the IPI program?

Analysis of the teachers' perceptions indicated that most pupils react very favourably to IPI mathematics and prefer working in the IPI system to working in a conventional program. It appears that when pupils have been exposed to an individualized program of studies, then the majority do not wish to return to a "whole class" approach. It also appears that involvement in the IPI system fosters pupil initiative, self-confidence and responsibility, and that behavior problems decrease. Pupil motivation is extremely high in IPI, and in the opinions of teachers, more mathematics is covered than in a conventional program.

Question Four: What perceptions do teachers have of the role of the aide in the IPI program?

The study revealed that IPI could not operate effectively without the assistance of teacher aides. It appears that the aides are essential for marking and other clerical duties so that the teacher can devote her time to helping individual pupils. The data also indicate that the aides have become an accepted and integral part of the schools involved in the IPI project.

Question Five: What perceptions do teachers have of the ways in which IPI has influenced the organization, administration and functioning of the school?

The results indicated that teachers appreciate the opportunity given them to individualize instruction, that they do not wish to return to the "traditional" methods of instruction and that IPI has a positive influence upon their educational philosophies. It is also suggested that IPI need not interfere with the general administration of the school and that it can influence the behavior of the administrators by bringing them into the classroom more than normally. Administrators, because of this contact, are in a position to "get to know the pupils better."

Question Six: What perceptions do teachers have of the manner in which IPI has affected the school system and the parental community?

The study suggested several extremely advantageous aspects of IPI are related to the parental community. It appears that IPI can assist in the improvement of communication with the parental community because the use of volunteer (parent) aides can bridge the gap between school and community, and because pupil enthusiasm for the program is transmitted to the parents. Secondly, involvement of parent aides can engender more

parental support from the community. Thirdly, the detailed IPI records give teachers a more concrete basis to discuss a pupil's progres with his parents.

In addition to the major questions, four subquestions were explored.

Sub-Question One: Do teachers in schools situated in an urban, higher socio-economic area; an urban, lower socio-economic area; and a rural area, have different perceptions of the IPI program in mathematics?

The study revealed that there were no major differences between the perceptions of the teachers in the three schools, (one rural, one in an urban, higher socio-economic area and one in an urban, lower socio-economic area). However, the data do reveal some minor differences. They suggest that pupil peer rivalry is more likely to increase in an urban school, and especially in one situated in a higher socio-economic area. Teachers from the rural school were also more prone than teachers from the two urban schools to consider that their role had become more professional, while teachers in the urban schools considered their role as a teacher had become more modern. The only school that perceived IPI provided well for the "below average" pupil was the school in the urban, lower socio-economic area.

Sub-Question Two: Do teachers with different teacher training backgrounds have varying perceptions of the IPI program?

The results indicated that the professional training a teacher has received does not appear to influence her performance in IPI or her perceptions of the majority of the aspects of the program. It is, however, suggested that the more highly trained a teacher is, the more conscious she becomes of the need for adequate conferences and planning sessions, and the more training a teacher has, the less likely she is to expect assistance from the teacher aides.

Sub-Question Three: Do teachers with varying years of teaching experience have different perceptions of the IPI program?

The study revealed that the less experienced teachers were more likely to request and need additional inservice conferences, planning and training sessions. The less experienced teachers considered that their role had become more modern since the introduction of IPI, while the more experienced teachers perceived that the assistance of aides had helped to positively change their attitude toward teaching.

Sub-Question Four: Do teachers' perceptions of the IPI program vary according to the grade or class taught?

The study revealed that pupil attitudes and behavior were more likely to change in the upper elementary grades than the primary grades after contact with IPI. It is also suggested that teachers of the upper elementary grades are not as prone as teachers of the primary grades to involve teacher aides in actual "professional" work, and that they are more likely to favour individualization in "all" subjects. The data also revealed that IPI provides well for the "below average" student in the upper elementary grades, while it appears to provide "unsatisfactorily" for the "below average" student in the primary grades.

In summary, Conners' study (1971) showed that the teachers were generally pleased with the success of the application of IPI, but had a number of important suggestions for future development.

Parental Opinions of IPI

As a final component in the evaluation, Hoke (1971) interviewed a random sample of 35 parents who had children in the IPI program. Three kinds of opinions were solicited: opinions of the program as administered in the school, effect of the program on children, and opinions of personal communication and administration details connected with the program.

A number of questions were asked of the parents concerning their thoughts about the IPI program itself. In response to a question asking about the desirable features of the program, the five most common responses were:

1. Children can work at their own speed.
2. Teacher aides were hired.
3. There were benefits using parent volunteers.
4. There were benefits in parent teacher conferences.
5. Instruction was individualized which developed the child's own scholastic and personality strengths.

When asked for the least desirable features of the program, the five most common responses were:

1. The slow unaggresive or lazy child was not pushed and given enough teacher supervision.
2. Parents, unable to volunteer were not involved and were frequently not informed about the arithmetic concepts completed by their children.
3. Children did not take arithmetic homework.
4. The desk work of the teacher prevented her from circulating around the class to teach, encourage, or supervise.
5. The IPI student profile sheet had little meaning for comparative purposes.

In response to questions concerning specific facts of the program, most parents were in favor of the children working at their own speed, but only about half of the parents were pleased with the lack of homework in IPI.

When responding to questions concerning the effect that the program was having on the children, about half of the parents thought that their child was achieving more than in the previous year, whereas only ten percent thought that the child was achieving less. Eighty percent of the parents were satisfied with the achievement. In response to questions concerning the effects of IPI on the children's future progress, some parents expressed concern with the lack of continuity between elementary and junior high school, however most of the parents indicated that experience with IPI would have a beneficial effect on the child's attitudes toward school in future years.

The final set of questions referred to the opinions of the personnel, communication and administrative variables connected with IPI. Most of

the parents were in favor of hiring teacher aides. They were also in favor of the use of parent volunteers in the classroom, although none of the parents interviewed was a volunteer. The child's progress was reported to the parents by means of an interview. Most parents liked these regularly scheduled conferences but a few criticisms were voiced: more interviews were needed; teachers can frighten parents if the parent does not have much self confidence; because the child is not at the interview, he cannot benefit from the feedback directly.

In response to a question asking whether or not the parents saw IPI as being worth the extra cost that was required, over 80 percent of the parents answered yes. Over 50 percent of the parents said that they would be in favor of an increase in taxes for improving educational standards using programs like IPI.

IPI arithmetic was valued for the effect it had of creating enthusiasm and motivating their children so that school became more interesting or likeable. Because of this, parents felt IPI was a superior instructional method, and was an improvement over the textbook approach. Several people felt the IPI method made the child's mind more alert and resulted in the learner being able to think more clearly. The concept of mastering materials was valued because parents felt that slow, medium, and fast working children would become more skilled in arithmetic. It was felt most children were benefitting from individualization of instruction; none were being held back or being forced to catch up with the class.

In general, there were a number of conditions which made the IPI project acceptable to the respondents. In the parents' estimation, the three experimental plants were excellent schools; parental communications with the school had been adequate during the first year of the study and also during the preceding year; the parents expressed respect for the qualifications, competency, and personalities of the school personnel; they were pleased with the quality of instruction in all subjects; and the parents felt that the various methods of reporting child progress complemented one another.

There were indications from parents that the IPI arithmetic program should have some restrictions placed on its implementation. Every parent interviewed, who had a first year (grade one) student working with the IPI materials, questioned the value of the instructional method for these students.

Summary and Discussion

In attempting to tie together the results of the evaluation, one becomes concerned that the particular schools chosen for the study may not be typical of the elementary schools in Alberta. To some extent this must be true. The IPI schools were chosen because of their willingness to innovate. The bases for nominating the Control schools were: the social class of the community served, size of staff, number of students, and similarity of physical facilities. It became apparent that the Control schools also possessed many of the same forward looking philosophies as the IPI schools.

In discussing the results it must be remembered that the participating teachers became very enthusiastic about the program and it is difficult to separate the effects of the program from the effects of the teacher enthusiasm. The results may generalize to other situations where IPI is used. We report what *can* happen not what will happen.

In capsule form, six conclusions can be highlighted from the research evidence.

1. IPI students did not achieve more than the Control students. In fact, where differences did occur, they favored the Control students. The achievement measures used, CTBS Arithmetic Skills and Problem Solving Subtests, are generally accepted as being among the most useful of the standardized tests of arithmetic achievement. Since they were designed to measure fairly traditional objectives, they were considered as appropriate in the present study.

2. The IPI students had a more positive attitude towards arithmetic than the Control students. The main difference between the two programs (other than the teacher variable) is that the children proceed at their own speed in the IPI program. As a result, there may be fewer frustrations. The sequence of material is about the same, in the two programs, so that time and the ability to work to a certain extent under conditions of their own making may be the important variables for student satisfaction.

3. The IPI teachers were more positively disposed towards innovation and individualization than their colleagues in the Control group. It seems unlikely that IPI *per se* had much to do with this finding. It is more likely that two situations acting in opposite directions combined to produce the result. Participation in an innovation—particularly one as highly structured as IPI—can easily produce a positive attitude toward the project. On the other hand, being asked to act in a Control group with no short term payoff can easily have the reverse effect on the Control teachers. These two factors in combination may account for the difference between the two groups. In any event, as a result of participating in the project, the teachers seem to have developed an openness to innovation. This openness coupled with the experience with a fairly successful program suggests that a pool of teachers now exists with whom curriculum development projects might be undertaken.

4. The IPI classroom allowed for more extended student interaction with teachers than did the Control schools. Often, one of the charges made against individualized programs is that they overlook the important social component of education. The results of this evaluation suggest that the reverse might be true. Extended teacher-student interactions are those interactions in which the student is allowed to talk to the teacher for sustained periods. This being the case, the results suggest that IPI students have longer conversations on an individual basis with their teachers than Control students do. A second interesting result of the interaction analysis was that IPI teachers used fewer controlling and criticising statements than their Control counterparts.

5. Parents of children in the IPI program were favorably impressed with the program and more than half of the parents would favor an increase in taxes to support the program. It is difficult to say if the parents' behavior would be consistent with their statements, nevertheless, their acceptance for all facets of the program was impressive. Many parents noted that changes had occurred in their children's attitudes towards arithmetic, and seemed to place a good deal of importance on this observation.

6. The parents freely gave their time to help in the many clerical chores that form part of the program. It is very difficult to say how the cooperation would be over an extended period of time. Clearly, the parents saw the program as experimental. It may be that as a program such as IPI became institutionalized, the parents would expect the boards to take over the responsibilities for clerical chores.

In discussing the results and attempting to reach some overall evaluation of the project, a number of factors must be weighed against each other, and to some extent each reader must pass judgment according to his own value system. The IPI program did not produce higher achievement on measures of educational objectives that are generally valued by society. The cost of the IPI program is greater than that of the standard elementary school arithmetic program. These two factors would suggest that the IPI program does not produce compensating benefits for higher costs. But the IPI program (and the teachers) do bring about more positive attitudes towards mathematics. How does this benefit stack up to the cost? Many parents felt that the results were worth more than the cost; however, one of the truths of educational research is that enthusiasm is an unstable commodity. Will the same results persist if the program is carried-on or instituted on a wider scale? Will parents continue to volunteer their services as the program becomes common place? Will achievement in the IPI schools actually fall behind that of the Control group as teacher enthusiasm is focused on other matters? Will achievement in the IPI schools become superior as teachers and students become more familiar with individualization? These are some of the questions that must be answered before decisions for adoption are taken.

References

Conners, R. D. An investigation of teachers' perceptions of individually prescribed instruction in mathematics. Chapter X in Oliva, F. D., Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971.

Hoke, M. Parental opinions of the pilot project in IPI arithmetic. Chapter XII in Oliva, F. D., Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971.

Holmes, R. Administrator's diary. Chapter IV in Oliva, F. D., Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971.

Maguire, T. O. Measurement of student attitudes. Chapter VII in Oliva, F. D. Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971.

Pacey, R. J. Project administrator's report. Chapter III in Oliva, F. D., Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971a.

Pacey, R. J. A comparison of classroom interaction in IPI and Control schools. Chapter V in Oliva, F. D., Maguire, T. O. and Pacey, R. J. IPI Project Report: Alberta Human Resources Research Council, Edmonton, 1971b.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 445-445*

Erwin Miklos

Reflections

The invitation to select an article for the 40th Anniversary issue of *AJER* did give me reason to pause and to reflect. My term as editor began with the production of the last issue of Volume 16. Consequently, the term falls well within the first half of *AJER*'s history.

My primary concern as editor related to ensuring the survival of *AJER* both financially and in terms of publishing quality articles. Although publication of the journal was supported in various ways by the Faculty of Education, the printing and distribution costs had to be covered by revenue from subscriptions and grants. In the early 1970s, production and distribution costs were increasing while grant and subscription revenue remained fairly steady. The problem of funding production on an issue-by-issue basis and using subscription revenue for future issues to cover current production was alleviated by the success with an application for a Canada Council publication grant in 1972. The grant greatly reduced the financial uncertainty at the time, and subsequent Canada Council and Social Science and Humanities Research Council support has continued to help keep *AJER* viable.

The awarding of a Canada Council grant, which involved a thorough external review process, also served to confirm that *AJER* had established a national reputation as an educational research journal.

With respect to maintaining and enhancing the quality of articles, I had hoped for but did not experience any dramatic increase in numbers of submissions. However, there continued to be an adequate supply from local, national, and international sources. I was impressed by the willingness of authors to revise manuscripts in accordance with suggestions of reviewers. Accordingly, I attempted to maintain the quality of manuscripts, in spite of a relatively low rejection rate, by providing a fair measure of advice to contributors for improving their manuscripts. I would like to believe that this service was especially important for the graduate students and researchers early in their careers who comprised a relatively high proportion of the contributors.

Although the cover design is not a highly significant aspect of a journal, I am pleased that the one introduced in March of 1970 has met with the approval of a succession of editors for the past 24 years. I hope that by mentioning this I am not raising any thoughts that the time has come to change the design of *AJER*'s cover!

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 446-447*

John McLeish
Editor, September 1972-June 1974

Native Education in the 1970s

Having been invited to select the outstanding article I chose to publish during my tenure of the post, I am in some difficulty not only because of the lapse of time but my failures of memory. It would be very easy to choose the worst, but I will leave readers to do that themselves. The problem is there are at least 10 articles that, on rereading Journal issues for these years, compete for the honor of being the best. Reflecting on my own experience of research articles in journals (as author, not as editor), I would single out A. Berger's piece on "The Education of Canadian Indians." This is subtitled "An In-Depth Study of Nine Families" (as it is, the reality counting for two merit points).

Without bafflegab, Berger points to the problem, hardly realized in the 1970s, that the dropout rate of Indian children in Canada was 94%, whereas with children in white schools it was only 12%. Noting that the situation was similar in the United States, Berger quotes Robert Kennedy (USA) as saying that this represents not only an national tragedy, but a national disgrace. This is all introductory matter, and excellent.

But the method, procedures, and data—not the opinions expressed—are the features defining quality research. Here these are specially commended. Nine families consisting of the fathers, mothers, and 40 children were interviewed as families on the Hobbema reserve near Edmonton. Four interviews, the first in English, the other three in the Cree language, were conducted by Georgina Trippe-de-Roche. No tapes were made of the family interviews, but the two interviewers wrote their recollections of what was said, these to be given to the family for use in the final exercise. The length of interviews was between one and seven hours. The attitudes of personal interest and concern expressed by the two interviewers easily won the involvement of the Indian families (the main impediment to such research). The authenticity of the study seized hold of the families involved to gain their cooperation. This is the major and outstanding feature of this study—its authentic flavor.

The finest part, however, nourishing this pervasive quality, were the procedures evaluating the statements made by the families. The first three interviews were broken down by content analysis into very brief single statements, which were typed on cards. These were presented to the families on the reserves with instructions to sort the cards into subject areas. The families selected between six and 39 such categories (e.g., cultural and tribal concerns, family, personal, religious, etc.). They were then asked to sort the cards into two piles: statements of immediate concern and those of less interest. The three most important, of immediate concern, were given as education, Indian heritage and culture, and family concerns. The fact is these Indian parents and children were very concerned about deficiencies in the administration and provision of education that

account for most of the Indian dropout. This article gave advance notice of Canadian Indian reactions of which we have become increasingly aware.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 448-456*

A. BERGER
The University of Alberta

The Education of Canadian Indians: An In-Depth Study of Nine Families

Education, heritage and culture are highly prized by families involved in this in-depth study of Canadian Indians. The study incorporates a number of relatively unique features; these include interviewing without using tape recorders, analyzing the data by means of content analysis procedures, involving the families in organizing the data, and writing the final report in the first person. Findings and conclusions contain implications for education and society. (Dr. Berger is Professor of Elementary Education, Faculty of Education, The University of Alberta.)

In Canada, the national dropout rate of Indian students between grades one and twelve is 94 percent; for non-Indian students, it is approximately 12 percent (Hawthorn, 1966). In the United States (U.S. Congress, 1968):

> Dropout rates are twice the national average; the level of formal education is half the national average; achievement levels are far below those of their white counterparts; and the Indian child falls progressively further behind the longer he stays in school.

In citing these and other statistics at the Hearings before the Special Subcommittee on Indian Education, the late U.S. Senator Robert F. Kennedy observed:

> These facts are the cold statistics which illuminate a national tragedy and a national disgrace. They demonstrate that the "first American" is in fact the last American in terms of employment, education, a decent income and the chance for a full and rewarding life (*Ibid.* p. 5.).

It is indeed unfortunate for Indians, as well as for the social and economic fabric of two great nations, that greater strides have not been taken to deal with this problem.

What are some reasons for the high dropout rate among Indian children? Can't they learn as well as white children? Don't Indian parents care about education? To find some useful answers, an interdisciplinary, in-depth study involving 40 children and nine families was conducted during the last half of 1971 and the greater portion of 1972.

Reprinted from *The Alberta Journal of Educational Research, 19*(4), 334-342, 1973.

The two-part study involved children in grades three and four in the Ermineskin School on the Hobbema Reserve. The permission of the reserve's Four Band Council was first obtained. Following a battery of individually-administered tests designed to assess the cognitive strengths of the children, an intervention program was developed on the basis of the test results. This produced significant gains in reading achievement in comparison with a control group of children. The children were using "simultaneous strategies" to solve problems requiring the use of "sequential strategies" (like reading), and *vice versa;* after the intervention program, their test scores increased dramatically.

The part of the study involving the families took place in or near their homes. Five of the families live in the City of Edmonton, one family about five miles west of the city on the Winterburn Reserve, and three about 50 miles south on the Hobbema Reserve. Each family was visited on four occasions. The conversations, which ranged from one to seven hours for each visit, were conducted either in Cree or English.

Before inviting these families to participate, discussions were held with representatives of the Indian Association of Alberta to determine the variables to be considered. The families were then studied with the following variables in mind: whether treaty or non-treaty; Indian or Métis; age of the children; mixed marriage or otherwise; varying levels of income.

The work with the families incorporated a number of relatively unique features. These included interviewing without using tape recorders; analyzing the data by means of content analysis procedures; involving the families in organizing the data; and writing a significant portion of the final report in the first person.

What follows here relates only to the design and rationale, analysis, selected findings and conclusions arising from the work with the nine families.

Design and Rationale

The general design of the study arose naturally from the question:

> What are some of the views and feelings of Indian parents in regard to education, culture, and related matters?

In attempting to answer this question, each family was visited on four separate occasions in or near their home. Information was collected during the first three visits and organized during the final visit.

With nearly every family, except where transportation facilities were not readily available the following pattern of visitations was followed:

Visits	*Visitors*	*Language Used*
First	Berger	English
Second	Berger and Trippe-de-Roche	English and Cree
Third	Trippe-de-Roche	English and Cree
Fourth	Berger and Trippe-de-Roche	English and Cree

Each visit lasted from one to seven hours. An attempt was made not to interrupt the ongoing lives of the families. On the first visit the families were informed that immediately following the interview, the recol-

lections would be recorded, transcribed, and returned to them, in written form, for them to organize on the final visit. To determine the accuracy of the interviewers' recollections, the following reliability exercise was conducted at the start of the study.

There was a conversation with two Indians not connected with the families. This was recorded. Immediately following this conversation interviewers' recollections of it were recorded onto another tape. Both tapes were then transcribed and the transcriptions were given to three independent observers to assess the quality of the *recollected* conversations against the criterion of the actual conversation as recorded. The results of the assessment made by these independent observers indicated a high reliability for this technique.

The conversations, or "interviews," were conducted in a manner suggested by Lewis Dexter (1970) in his book, *Elite and Specialized Interviewing,* and by Alfred Benjamin (1968) who, in *The Helping Interview,* writes:

> I do not see the interviewer as passive in the least. On the contrary, I see him as active at all times. I am not implying that he should talk a great deal, but I am saying that he should make his presence and interest continuously felt. The interviewer is active in gaining as deep an understanding as possible of the interviewee's world. . . . At all times he is active in revealing himself to be a person deeply involved in another person.

A rationale for in-depth studies is provided by David Fox (1969) in *The Research Process in Education.*

> There are mass surveys as to the sociopsychological characteristics of children who ultimately become anti-social as to be legally classified as juvenile delinquents. We know that most often they come from broken homes, that, except for sex offenses, they are more likely to be boys than girls, that their social malfunctioning is accompanied, and usually preceded, by academic malfunctioning, and so on. Yet we still have not learned from this mass survey approach how the different deprivations and difficulties interact so that some children become serious delinquents and others, who exhibit the same characteristics identified by the mass survey, do not. The answer lies in other characteristics yet to be identified, and/or in the pattern of interaction of the characteristics in the individual. It is at this level that the case study approach would function, studying the characteristics as they exist in company with each other within the person and life space of individuals.

Further rationale can be found in the preamble to *Let Us Now Praise Famous Men* (Agee and Evans, 1941) and in the Introduction to *La Vida* (Lewis, 1966).

Content Analysis

On the fourth visit each family engaged in a series of intriguing sorting processes to organize the data obtained during the first three visits. The information obtained during the conversations with each family had been previously subjected to content analysis. The typed recollections were scrutinized to cull each meaningful unit of expression (e.g., Harold Cardinal is a great man). Each of these units was typed on a 3 x 5 card; each family had up to 75 cards. On the fourth visit, then, the family placed each of their own cards into categories of their own creation. That is, they looked at a card, said it was "religion" (for example): the

card was then placed in an envelope on which the word "religion" was written. At the end, some envelopes contained many more cards than did other envelopes; some families had many more envelopes, representing categories, than other families.

The families were engaged in a final sorting process. They took the envelopes bearing the categories and made two piles, one containing items of immediate interest and concern, and the other containing items of lesser interest and concern to them.

With the permission of the families involved, the cards that had been made into the fewest number of categories (six) were exchanged with the cards that had been made into the largest number of categories (39). The intent was to see how many categories each family would make with the other family's data. Family D made 39 categories with their own data and 37 with Family C's data; Family C made six categories with their own data and 10 categories with Family D's data. This result was expected,

TABLE 1

TOPICS OF IMMEDIATE INTEREST AND CONCERN TO THE FAMILIES

Common Categories	Families									Totals	
	A	B	C	D	E	F	G	H	I	**	*
Education	**	**	**	*	**		**	**	**	7	1
Heritage and Culture	**	**	**	**	**		**	**	**	8	
Family	*		**	**	**		**		**	5	1
Indian Organizations	**		**	*					*	2	2
Discrimination	*	**	**	*					*	2	3
Integration			**	**	*			*	**	3	2
Employment	**	*		*	*			*	**	2	4
Treaty Rights	**	**		*	*					2	2
Personal Concerns	*			*				**	**	2	2
Language		**					**		**	3	
Religion	*			*	*		**			1	3
Social Status				*				*	**	1	2
History	*			**						1	1
Communication	*						**			1	1

** Indicates "of immediate interest and concern"
* Indicates "of lesser interest and concern"

[1]Family F did not continue beyond the first conversation for reasons of a personal nature.

[2]Extrapolated from cards and categories made by Family I.

the assumption being that the processing of information is not largely affected by whose information is being processed.

The two interviewers scrutinized the typed recollections to see what was actually discussed during the conversational visits. This was for the purpose of determining whether the families chose to discuss certain topics with the Indian visitor and other topics with the white visitor. Two people also (independently) examined all of the cards in all of the envelopes to form common categories.

Selected Findings

The upshot of the organized data appears in the table showing common categories of immediate interest and concern to the families. Of most immediate interest and concern for most families are education, heritage and culture, and family.

Conclusions

As might be expected a great number of surprising insights were achieved as a result of these visits and conversations. Indians, as other people do, express concern with many of the same problems with which we are concerned. Contrary to the stereotype of the Indian being unconcerned about education, the families in this study talked more about this than about any other topic. They saw education as the way to a better life for their children, and desperately wanted them to succeed in school. One family was distressed about their son who had dropped out of school. The only family that indicated "education" to be of "lesser interest and concern" was one where all the children were grown-up and employed.

In conversations with the families and others, it also appeared that Indian people have special problems that directly affect their education. The boy who dropped out of school is illustrative of the many Indian boys who must leave the reserve after grade nine if they want to continue their education. The fact is that most reserve schools do not go beyond grade nine; many do not go beyond grade six. What this means is that these young people must leave their families and go off by themselves, live in a room in the city, and continue their education. It is a trying experience for those who are unprepared for city life. They may be told which buses to take to get to their school, and how to return, but it is often overlooked that it is also vital information for reservation-reared children to know about how to get on and off the bus, or how to transfer to another bus. Some of the families indicated that their children had found a room with old couples who needed the income but who also needed a degree of quiet that few normal teenagers are able to provide. Such problems often arise in regard to the homes in which these youngsters are placed. Often overlooked, also, is the fact that many Indians are very poor, sometimes owning only one pair of clothes. As a result, some children who go to school in cities or smaller communities have to miss school on wash day.

Many Indian people know these problems, but are not in a position to influence changes. This is because they have very little control over their own education. In the Province of Alberta, which provides some of the best education on this continent thanks to the province's wealth

from oil and natural gas resources, it is not legally possible for Indians to sit on Boards of Education. This is information given by a spokesman for the Department of Indian Affairs.

While most reserves have school committees, many are largely ineffective, with Indians having little or no say about who should teach their children. As a result, one Indian family living on the Hobbema Reserve within sight of Ermineskin School sends their children to school in a nearby "white" community because, the husband explained, many of the children cannot understand some of the foreign-born teachers who speak broken English.

If Indians were permitted a little more involvement in their children's education, they would be better able to point out the need for a useful orientation program to help prepare children who must continue their education in the cities. They could also encourage teachers to visit the homes of their school children.

> . . . teachers would be welcomed in many homes but the Indian people are shy and so the teachers must make the first move. The teachers should take the initiative in expressing interest in visiting homes. She mentioned one teacher who sent notes home expressing his interest. The children gave the notes to the parents and some of the parents invited him to come and visit with them.
>
> —Family H (Hobbema Reserve)

During a classroom discussion about teachers visiting the homes of school children, one teacher working in a northern Indian community said that teachers there tried to visit the homes but the parents refused them. On further questioning, however, it turned out that a carload of teachers came to visit each of the Indian homes. Most white parents would probably be filled with trepidation at the sight of a group of teachers piling out of a car parked in front of their home, unannounced.

Teachers visiting homes—and administrators giving them time to do so—is a controversial idea. This was made clear to me this summer by an Indian woman who had completed an orientation course for teachers of Indian children. She claimed that students were told *not* to visit the homes of Indians because of the possibility to talking about "politics." The fact remains that the "wrong" parents—in the sense that those who come are relatively well informed—tend to go to parent-teacher meetings. Those who do not come are those who, for their children's sakes, need to be reached most urgently. In trying to reach them, it is helpful to remember that many people feel more comfortable in the familiar surroundings of their own homes rather than within the confines of an imposing school building.

The families were relatively unprejudiced about ethnic differences as between teachers and taught.

> I asked her if it is necessary for the teacher of Indian children to be Indian. She felt that the main thing is that the teachers should be good people. The topic of teacher aides came up and she said that it was a good idea to have Indian mothers as teacher aides.
>
> —Family B (Edmonton)

What was striking about Indians along this line is their openness to

share, to accept white people, if only they—the whites—would make the first move.

What happens in most communities is that teachers tend to remain relatively ignorant about the daily lives of the parents and children with whom they are involved. On some reserves, teachers actually cluster together in motel units, with only the braver ones daring to venture out into the culture and the homelife of the people.

The influence of language on thinking and behavior is no secret. As long ago as the 1930s Rudolf Carnap (1937) and B. L. Whorf (1956) studied the influence of language, with the latter observing some of its effects on Hopi Indians. In our study, the effects of language along with other factors were strikingly evident when two families processed each other's data as well as their own. It is very important that fluency and flexibility of language behavior be encouraged.

> Mrs. G. said that TV has helped her children; they pick up English and learn about other people as well.
>
> —Family G (Hobbema Reserve)

During the year of the study I learned that I had some stereotypes about Indians—and that some of these stereotypes had been reinforced by Indians themselves! When I went to hear Kahn Tineta Horn (an outspoken Indian woman, who several years back won the title of Princess in the Canadian Annual Indian Princess Pageant) speak at The University of Alberta, she told of her brothers who were high steel workers, and she indicated that Indian children must have an education that is connected with the outdoors. I was startled, in discussing my findings with Indians and other people, how I also had associated Indians with the outdoor life, completely neglecting to take into consideration that *very few people*—Indians *or* whites—want to be "tied down" to a desk for eight hours a day.

While it may be stereotyped thinking to educate Indians solely for the outdoors, curriculum makers should not overlook the harmonious relationship which Indians tend to enjoy with nature, a relationship which is reflected in much of their talk and writing as well as in their lives. To overlook this relationship is to overlook a vast wealth of resources—to the detriment of Indians and whites alike.

There is much interest now in knowing more about the past and present lives of Indian people.

> Mrs. A commented on her schooling and said they prayed fourteen times a day. I thought this was rather remarkable. She indicated that they prayed when they got up. They had mass before breakfast, they prayed after breakfast and on through the day. They also prayed for the Pope and they prayed for the Communists. She mentioned an arithmetic teacher who had come from Toronto who, she said, "was a real fanatic". Every so often during the day, after they had finished their arithmetic, they would have to get on the floor and say some prayers. She said they also prayed for being sinners. She said she doesn't know how they could possibly have sinned except for the possible sin of everyone thinking they should kill the head nun.
>
> —Family A (Edmonton)

It is good for people to know their own heritage and culture as well as the heritage and culture of others. But the argument is sometimes put forward that, by knowing about one's ancestors and customs, one be-

comes a better person. There may be some truth in this argument. But care must be taken that it does not cloud the fact that people feel good when they accomplish things successfully.

On an airplane trip last summer from New York to Buenos Aires, I sat next to a young man who was born in Italy, grew up in Argentina, then came to New Jersey where, through dint of hard work, he managed to save enough money to open a restaurant; a few years later, at the time of our conversation; he owned a successful restaurant and a house with a swimming pool in New Jersey and a small hotel in Italy. He observed, in broken English, that while he is not able to do many things, there is one thing that he can do better than anyone else in the world—and that is, make the very best pizzas. It seems to me that this is the attitude that must be fostered in schools. Each child must feel, within the realms of his own culture, that he can do something very well. Some may challenge this suggestion on the grounds of a lack of competition among Indians, but what may appear to be a lack of competition is, in reality, politeness on the part of Indians. (It is simply not good manners to show you know answers when others may not know them.) Polite behavior and successful accomplishments can go hand in hand in any culture.

Knowledge of one's own roots is vital, but the fact remains that young people must be able to perform well in school. Knowing one's own culture, while comforting, is not much help to the child who cannot read his books adequately. The books used in most classrooms on this continent are difficult for most children, and the educational problems are aggravated when there are additional potential hurdles like bilingualism and living away from home.

It would be educationally-beneficial for an interdisciplinary team to examine the data derived from this unique in-depth study of nine families and 40 children and then, after interviews with the *children of the families,* develop a meaningful language arts curriculum followed by an in-depth training program for teachers of Indian children.

There are still many questions to be answered; some are vitally relevant to the Indians. One question that still puzzles me is why Indians gave up their language and their religion so readily. True, many now have two marriage ceremonies—one of which is Indian—but most know little about their precious religious heritage, and if I were an Indian I would like to know what happened.

What is also surprising is why Indian leaders have not engaged in a more fruitful public relations program with nearby white communities. In some communities small groups of white children go to school on the reserve nearby, and *vice versa*. But the interaction is very limited. I recall an incident that happened when I drove for the first time alone to the Hobbema Reserve. I stopped at a gasoline station in Wetaskiwin, about ten miles away from Hobbema, filled up, and checked my directions. The teenage white boy was startled: "You're not going to the reserve by yourself, are you?" He told me he had never been to the reserve. That evening, while at the powwow ceremony. I thought of the young man and how much he was missing, even though he lived only ten miles away. In a very real sense *he* was culturally deprived.

There are questions in regard to the psychological impact of the *Indian Act* which have never been properly explored. It is good that Indians receive recompense and reparation for what happened years ago. But what does this inherent paternalism do to the initiative of Indians? Perhaps nothing, but it certainly seems a question worth exploration. For in a very real sense we are all displaced persons. Who among us has not had his roots torn, as if by a shudder of the earth? Who has been able to "hold fast to dreams"? Who has not experienced a sorrow, a loss, and a longing for the past? We are an uprooted people on this planet, and unless we can break the shackles of the past while extracting strength from our heritage we are all of us a lost people hurtling through space.

Acknowledgments

The following are among those whose contributions to the study of nine families should be acknowledged: Georgina Trippe-de-Roche, Linda Halfe and Judy Half Casteel, Vicki Crowchild and Jenny Margetts, Chief Maurice Wolfe of the Ermineskin Band, the Hobbema Four Band Council, and Joseph Couture of the Alberta Indian Education Center. Appreciation is also extended to Les Gue (Department of Educational Administration), Bob Jackson (Department of Elementary Education), J. P. Das and Tom Maguire (both of the Department of Educational Psychology), and Myer Horowitz (Dean of the Faculty of Education and former Chairman of the Department of Elementary Education). The research reported herein was performed pursuant to a contract with the Department of Indian Affairs and Northern Development (Ottawa). The Final Report is available on microfiche through ERIC ED 071 812. Parts of this article appeared in *Elements* and the *Journal of American Indian Education.*

References

Agee, James and Walker Evans. *Let Us Now Praise Famous Men.* New York: Houghton Mifflin, 1941. (Reprint by Ballantine Books, 1966, pp. 7-15).

Benjamin, Alfred. *The Helping Interview.* Boston: Houghton Mifflin, 1968, p. 38.

Carnap, Rudolf. *The Logical Syntax of Language.* New York: Humanities Press, 1937.

Dexter, Lewis Anthony. *Elite and Specialized Interviewing.* Evanston, Illinois: Northwestern University Press, 1970.

Fox, David J. *The Research Process in Education.* New York: Holt, Rinehart and Winston, 1969, p. 427.

Hawthorn, H. B. *A Survey of the Contemporary Indians of Canada: A Report on Economic, Political, Educational Needs and Policies.* Ottawa: Department of Indian Affairs and Northern Development, 1966 (volume one, 409 pp.) and 1967 (volume two, 251 pp.).

Hearings Before the Special Subcommittee on Indian Education and Public Welfare: Ninetieth Congress of the U.S. Senate: First and Second Sessions on the Study of the Education of Indian Children. Published in Five Parts, 1967-68. (Available from the U.S. Government Printing Office, Washington, D.C.).

Lewis, Oscar. *La Vida.* New York: Vintage Books, 1966, pp. xi-lv.

Whorf, R. L. "The Relation of Habitual Thought and Behavior to Language," in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* (John B. Carroll, editor). Cambridge: MIT Press, 1956.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 457-457*

Henry W. Hodysh
Editor 1977-1981

Hersom and the Development of Educational Research

It is difficult to identify a single article published in *AJER* during my tenure as editor that might be considered superior to the rest. Of the many that were printed each in its own way is a valued contribution to scholarship.

Naomi Hersom's "Twenty-five Years of Research in Education," a "Perspectives" essay, is of particular interest, for it provides a reflective analysis of *AJER*'s influence on the growth of educational research in Canada (Vol. 26, No. 4, December 1980, pp. 262-275). Mapping the journal's shifting emphasis from its appearance in 1955, she presents a synthesis of the central elements shaping the nature of research and examines the descriptive, analytic, and theory-oriented approaches that at various times dominated its interpretation.

Hersom's reminder that social and intellectual forces play a critical role in educational research is especially significant. Her observations neatly focus a vast and complex body of ideas, including, for example, the relationship between basic and applied studies, a question of increasing importance in today's rapidly changing educational scene. Other issues explored include the nature of theory and practice, the need for new approaches to the investigation of educational questions, and the challenges that would confront the research community in the future.

In keeping with *AJER*'s mandate to critique and disseminate ideas, Hersom identifies potential gaps in the creation of knowledge and, in the light of the journal's historical development, suggests alternative directions for better understanding the student, curriculum, and research at all levels of education.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 458-471*

PERSPECTIVES

Twenty-five Years of Research in Education: *The Alberta Journal of Educational Research* 1955 to 1979

Naomi Hersom
The University of British Columbia

To search more diligently and to be driven on by curiosity is commonly the lot of one who chooses to become a researcher. Prolonged contemplation of a problem and the periods when progress seems lacking or painfully slow are frequently prerequisites to that splendid moment of perception when something is seen to be true about the world. But the private moment belonging to the researcher alone is not sufficient. It must be shared, made public, opened up to the tests of others' perceptions and ideas, too. It is for this reason that the publication of the results of research is so vital—publication represents a formal invitation to criticism. And for a quarter of a century, *The Alberta Journal of Educational Research* has been one of the few publications in Canada which has afforded an opportunity for researchers in education to put forward the results of their efforts for careful scrutiny and for practical application.

It is intriguing to contemplate the interplay of events and ideas shaping the nature and direction of research activities in education primarily, though not exclusively, in the English-speaking world. The articles selected for publication in

Reprinted from *The Alberta Journal of Educational Research*, 26(4), 262-275, 1980.

AJER over its twenty-five year history provide a rather unique opportunity to do just that. While its editorial policies have developed and changed somewhat, the overall goal of presenting ideas that represent the best in all aspects of educational research has not. Accordingly, one can find in its pages examples of developmental trends in educational research that transcend international boundaries while at the same time demonstrating the ways such trends influence research being done in Canada. More specifically, *AJER* has reflected the research activity to be found in a university committed to fostering excellence in the study of education (Swift, 1970). At the outset of this retrospective look at the *Journal*, it should be acknowledged that the University of Alberta has demonstrated the seriousness of that commitment by its longstanding support of this publication. The *Journal* articles, seen as a whole, provide a mirror image of the ebb and flow of interests and emphases, the growth and decline of particular ideas, and the effects of change in society at large on the questions and problems addressed by researchers in education.

Educational research in North America generally has changed in style and scope within the context of demographic and other social changes surrounding the whole of the academic research enterprise. In Canada, we cannot set aside the extensive influence of public policies and government resource allocations for educational research in the United States which uninhibited communication of ideas and the extensive exchange of researchers has induced. One can argue that the pursuit of knowledge knows no bounds, but at the same time it would be foolish to discount the effects of the larger environment on the quality and quantity of research effort. With this presupposition in mind, an examination of the articles which have appeared in *AJER* suggests that there are at least three main sources of influence. Besides the national and international events, policies, public attitudes—the general milieu in which researchers do research in education—there are the ways in which researchers interact with one another and with the professional and public monitoring of their activities. More specifically, the *AJER* articles may be viewed as a fairly faithful reflection of the ways in which concepts about research in education have developed over a particular period: notions about what counts for valid knowledge, the appropriateness of various techniques and technologies, the interplay between the world of ideas in the social sciences and humanities and in education, the dissemination of research findings, and the effects, if any, of their application on practice.

The ecology of educational research in Canada

An ecological view of educational research by definition places emphasis on development in relation to total environment. Research in education, like research in the social sciences generally, grew rapidly following the Second World War. While social scientists claimed that their research would provide an understanding of humankind in general, which in turn would lead to the formulation of solutions to social problems, researchers in education focused on greater understanding of people in schools in order to formulate solutions to educational problems. Educational researchers in universities shared with their counterparts the conviction that analyses of processes and mechanisms would provide the knowledge required to resolve problems and to invent better policies and

practices. Governments viewed education as an economic investment by society that would produce better individuals, intellectually and personally, who in turn would be better citizens. Funding for research was increased both to help resolve social problems through education and to contribute to the quality of preparation for productive work and thought in Canadian society. Thus it was that educational research activity changed dramatically by the mid 1950's. It was no longer seen to be merely an activity to be carried out by professors of education during time not devoted to teaching: instead it was thought to be an instrument to be used to cope with the present and to influence the future.

At the same time, international events took place, heightening awareness of philosophical and scientific thought in education elsewhere, particularly in the Soviet Union. Reports and studies of differences in theories and practices and their resulting effects became sources of ideas for possible improvements in Canadian education. Later, when radical changes in American equity laws were so painfully being born, the course of educational research began to take on new directions. A preoccupation with ways to predict group achievement and to standardize treatments of groups began to give ground to a search for ways to provide equality of opportunity through education which would offset the effects of poverty and discrimination and would take into account the characteristics of individuals.

Thus one finds in the earlier volumes of *AJER* an article by Sparby (1959) addressing the relationship of research in education to the social sciences and the value of using an interdisciplinary approach, succeeded by a series of reports of research studies which depended heavily on social role theory to examine the behaviours of students, teachers, and administrators. The Greenfield and Andrews (1961) and Keeler and Andrews (1963) reports are typical of those studies undertaken with the specific intent of conducting educational research in order to improve the quality of schooling. The focus at the time was on the effectiveness of actors in fulfilling designated roles in educational organizations, but five years later, one of those authors (Greenfield, 1968) reassessed the use of the concept of role behaviour in educational theory and research and suggested new directions for future research. Or one can find, beginning with MacDonald (1959) a continuing interest in Soviet education: Stewin and others (1969, 1974) investigated the Soviet theory of set; Smith (1973) looked at Soviet studies in language acquisition; and Gulutsan (1967) was concerned with the relationship between the ideas of Piaget and Soviet psychologists. Similarly, a change over time appears in articles about tests and testing. In 1956, Clarke and Dunlop wrote about "Tests in Measurement and Evaluation" and studies on the use of standardized tests of achievement such as Sangster's (1956) gave way eventually to research growing out of the criticism of testing related to such factors as the effects of culture (Safran, 1963), anxiety (Frost, 1965), creativity (Cropley, 1967), and divergent thinking (Vernon, 1972). Gradually the ideas of Jean Piaget became more and more prominent in the studies appearing in *AJER*: in 1960, Crawford reported a study of Piaget's thought in relation to teaching school mathematics; in 1967, Côté's study of Piaget's formulations on perception and intelligence appeared; by 1973, O'Bryan and MacArthur had made an analysis of Piaget's notion of reversibility and Lefrançois had investigated Piaget's model of equilibrium through adaptation.

It can of course be argued that these themes appear in the pages of *AJER* largely because they reflect the choices of the referees and the editors of the day. To some extent that may be the case, but that is too simple an explanation. Final selections by editors are made from a pool of articles available, and that pool is the result of the research activity of people who have pursued ideas or who may have been supported in their pursuit of those ideas for a variety of reasons. It is not surprising that some of the research done in education and some of the motivations of researchers in education are related to social and political factors much larger than the educational enterprise itself. Sources of funding and demands for research in education which come from outside of the universities illustrate this point. The effects of such interventions are most apparent in the work in educational administration associated with the University of Alberta.

In the preface to Swift's (1970) volume entitled *Educational Administration in Canada,* Stewart describes the people who were involved in the process of establishing a special centre for studies in school administration at the University of Alberta. He cites the following as influential factors: an expressed concern among members of the Canadian Education Association in the first decade after the War that more facilities for post-graduate study in education be made available in this country; a disposition among Deans of Education in Canada to cooperate to that end and their attitude of good will toward the possibility of development at the University of Alberta under the leadership of H. T. Coutts, Dean of the Faculty of Education; the support of the Alberta Minister of Education, the University Board of Governors, the President of the CEA (in office during 1954-55); and Dr. Arthur Reeves, the person "who brought the scheme to fruition" by preparing and promoting the proposal for the CEA-Kellogg Project (pp. *iii, iv*). The W. K. Kellogg Foundation offered support for an action program

> which would have an immediate and beneficial effect on the educational process and, perhaps more important, which would serve to set in motion programs which, having demonstrated usefulness, would be taken over, and be financially supported on a continuing basis by Canadian school authorities and educational institutions. (pp. 13-14)

Subsequently, as Swift suggested, a program of research was undertaken which was unlike the kind of research "as that term was generally understood." Examples of some of the studies completed in large part because of the mandate given to the Department of Educational Administration at the University of Alberta, funded by the Kellogg Foundation initially, and later by the University's own resources, are to be found throughout the pages of *AJER*. The articles provide some evidence that the concerns of practitioners, as voiced by the Canadian Education Association, and the concerns of informed members of the larger public, as represented by the Kellogg Foundation, had an effect on the course of research in education. In this case the interdependence among groups was overtly welcomed by Dean Coutts and the Faculty of Education, and the outcomes have been documented in considerable detail by Swift. One suspects that in other areas of educational research, "outside" influences have not been recognized as clearly but are nevertheless no less active in shaping educational research efforts.

The influence of the scholarly community

Another of the factors shaping the nature and direction of educational research is almost certainly the type of program designed to prepare researchers (Willson,

1980). The majority of the programs offered in North America twenty-five years ago were embedded in the requirements and traditions of university faculties of graduate studies. The practice continues to a great extent today. The overall result is that the ideas of the scholar-models in universities who supervise theses and who are members of thesis committees have a long-term influence on the types of problems addressed by researchers in education and on the means of exploring them. With the growth of graduate programs in the faculties of education in Canada, more professors and an increasing number of graduate students carried on this academic tradition and, quite naturally, shared beliefs and assumptions about the research enterprise. It also seems reasonable to suppose that a centre like the Division of Educational Research at the University of Alberta, in the process of facilitating research activity, has been instrumental in reinforcing ideas and developing standards of excellence in educational research.

In some cases the traditions are longstanding and deeply rooted, such as in the field of testing and test development, and researchers have consistently addressed certain basic issues over time. In 1956, *AJER* published Powell's evaluation of the "Detroit Beginning First-Grade I.Q. Test" in terms of its validity, reliability, and standardization. In 1974, an article by Randhawa, Hunt and Rawlyk reported a study which included, among other things, an examination of the reliability and validity of the "The Canadian Cognitive Abilities Test." These and other studies focusing on the technical aspects of perfecting tools-to-be-used share certain characteristics handed down through a long line of researchers. Walker's article in the 1959 volume, assessing the influence of Binet after half a century had elapsed, was possible partly because certain presuppositions about the nature of learning and testing remained relatively constant among groups of researchers. This is not to imply that the field remained static, but that there had been a certain kind of continuity. Changes were introduced when some researchers became concerned with studying the effects of tests-in-use (Frost, 1965), or when the need for tests based on different kinds of assumptions about the nature of learning and thinking came to be held by the scholarly community (Hobbs, 1975). But, overall, it can be said that those scholars in education who see themselves primarily as researchers concerned about the application of the methods of social science research to the problems of education have actively sought to improve the quality of their research and to strengthen programs for training researchers.

What may be seen to apply to the influence of the whole community of scholars in education nationally and internationally may also be seen to affect the research activities of groups of scholars who share an interest in a particular subject area or, on a small scale, among scholars located in a single institution. When fairly stable associations are maintained over the years, expectations for research are built up which are productive and influential for students and colleagues alike. One sizeable group of articles published by *AJER* seems to illustrate this phenomenon. Beginning in 1956 and continuing regularly throughout the pages of *AJER*, there are reports of studies on school mathematics programs, the preparation of teachers of mathematics, and the performance of students doing various types of mathematics tasks. The researchers whose names are associated with some of these studies seem to indicate the effects of

longstanding associations among researchers in mathematics education: Nelson and Worth, 1960; Nelson, 1964; Nelson and Towler, 1967; Liedtke and Nelson, 1968; Bana and Nelson, 1977; Bourgeois and Nelson, 1977; Nelson and Kieren, 1977; and Kieren and Nelson, 1978.

Another facet of this phenomenon of commonality in research interests and the effects of leadership on research directions seems to be apparent in the body of work published by *AJER* related to Soviet scholarship. The heritage of western Canada in particular has been enriched by its Eastern European roots and by those Canadian scholars who have maintained communications with European scholars, specifically with their Soviet counterparts. At the University of Alberta, Gulutsan (1967) was a founding member of the interdisciplinary Centre for Eastern European Studies and it is noticeable that several of his colleagues in the Department of Educational Psychology, coming themselves from a variety of research traditions, produced a number of studies analyzing various aspects of education in the Soviet Union. MacDonald (1960) wrote "Explanation and the Neo-Marxist Theory of Mental Testing" at a time when the race to explore space was reaching its competitive heights. A decade later, Hritzuk and Janzen (1970) studied the concept of set held by two Soviet psychologists, Luchin and Uznadze, followed by Lefrançois' (1973) account of behaviour theory, and by Cowper and Stewin's (1974) experimental critique of Soviet set theory. Other psychologists, like McLeish (1974), examined certain features of Soviet education from a social system perspective or, like Stewin and Martin (1974), from a developmental perspective. It would be dangerous to ascribe disproportionate effects to the influence of scholars on one another, but it does seem reasonable to identify personal interests and group contributions among those factors which, for whatever sets of complex reasons, have led researchers to follow one course of inquiry rather than another.

This same factor seems to be present in part in the process of change which can be seen in the research on teaching and teacher preparation during the past twenty-five years. At a time of teacher shortage in North America, Clarke and Pilkington (1955) and Aikenhead (1956) investigated the reasons why students and practising teachers were not choosing to enter or to remain in the profession. Their studies used questionnaires designed to assess the opinions and attitudes held by teachers and others. In the next decade the focus of research shifted to teacher performance in the classroom and attempted to identify the kinds of knowledge and behaviours associated with successful teaching. So, for example, Eddy (1963) studied "Teacher Characteristics and Social Studies Achievement" and McBride (1964) investigated student assessments of teacher performance based on the assumption that when students are generally satisfied with their experiences in a classroom they will derive greater benefit from the teaching and the learning experiences offered to them. By the mid-1970's, the notion of the effect of teachers' attitudes towards pupils and teacher-held expectations for pupil performance became predominant in the research. Lavoie and Adams (1974) conducted a field study to ascertain the ways teacher expectations may be affected by children's physical attributes, sex, and conduct. Hoxter (1974) investigated teacher attitudes to find out how these affect students who come from cultures other than the teacher's own. Miller (1974) studied changes in student teacher

attitudes towards children and schooling, reflecting the changes in values held by students throughout the period of preservice professional education. Recent research has continued the emphasis on the internal rather than the external factors related to success in teaching. Where the work of Clarke and Pilkington (1955) emphasized the effects of public opinion and of home and school influences on prospective teachers, Noll, Willower and Barnette (1977) directed attention to the teacher as a person by using measures of "self-actualization" and measures of attitudes in terms of "pupil control ideology" to predict the nature of teaching behaviours, and Cheong and Wadden (1978) used measures of teacher attitudes to predict effects on pupil self-concept.

If my initial presupposition holds true, that *AJER* can be seen as a sort of weathercock for directions in the developments taking place in educational research in North America, then the articles to be found in its twenty-five volumes to date reveal something of the state of the art in Canadian research on teaching and teacher education. Although one would expect to find many educational researchers making the investigation of teaching and teacher preparation programs central to their work, especially since so many are conducting their studies in the context of professional faculties, this does not seem to be the case. The studies appear to be so diverse that it is difficult to find ways in which research findings could contribute over time to the building of an adequate body of knowledge. Some possible explanations for this state of affairs seem to emerge from an overall examination of the reports of research appearing in *AJER*. One observes that there is an almost exclusive association of researchers within a discipline or a field of study but little association across disciplines. While inquiries by researchers in educational philosophy, sociology, psychology, and history or administration and curriculum are ongoing, few of the studies have been undertaken in an attempt to work across discipline lines or to synthesize findings in order to discover general principles or theories. Work like Lindstedt's (1960) study on "Teacher Qualifications and Grade IX Mathematics" identified certain features characteristic of preservice teacher education and subsequent teaching experience in schools that can affect the levels of competence attained by teachers of mathematics. Greenfield and Andrews (1961) identified certain types of teacher behaviours that can be associated with pupil achievement on school examinations. There seems to be little evidence of concern on the part of the community of educational researchers to take the findings of such studies consciously into account when planning subsequent investigations. Even though a researcher like Poole (1975) set out to find ways to improve teacher education programs by identifying component skills of teaching and by deliberately training people to generate ideas on the assumption that this would make for more creative teaching in the classroom, the study was conducted on a relatively small scale without benefit of cross-consultation with other researchers such as I have envisaged. Of course, one might argue that a great deal of the research conducted in education implies the need to know more about teaching and teacher preparation and thus contributes to our knowledge indirectly. But, historically, the resolution of difficult problems in any field of inquiry has not come about by circumventing them. Obviously, different researchers will use different approaches in searching for new knowledge and may even be in sharp disagreement with one another in their views of the best approaches to take, but they share in a common

pursuit and the influence of the community of scholars makes itself felt. In the community of scholars in education there appears to be lacking a continuing, persistent commitment to the kind of research that boldly identifies itself with the intent to pursue new ideas, to invent new practices, to resolve longstanding problems which would lead to the improvement of university and field programs in teacher education.

The traditions of the scholarly community can be tremendously productive but there is a danger that those same traditions will become paternalistic and conformist in their effects on research. The apprentice researcher who is introduced to a particular mode of conceptualizing problems-to-be-solved in education may find it difficult to accept or to explore other ways of thinking and to use alternative modes and techniques of investigation. The myth of "the scientific method" has influenced research in education in North America as it has in other fields of inquiry and has played a predominant role in the research in education reported in *AJER*.

The extension of knowledge in education through research

To this point the *AJER* has been viewed partly as a product of its time and place and partly as an illustration of the several ways the community of scholars in education shapes the work that is ongoing in the field of inquiry. Undergirding both perspectives are certain basic notions about what educational research is and what counts for valid knowledge in education. Is research in education basic or applied and, if it is thought to be applied research, does that mean it is applicable to further academic study or to practice, or to both? Baxter (1980) argues that research in the applied sciences is "basic in the sense that it is concerned with knowledge and methods that might be applicable in a number of different situations" and thus it enjoys a kind of strength and versatility which would not be possible if it were highly specific (p. 3). Others would draw the line more clearly between basic and applied research. Mattersich (1980) distinguishes basic from applied research by emphasizing its intents: when research is directed toward increasing knowledge it can be termed basic; when it is directed toward practical applications of knowledge, means-end relationships, or the resolution of specific problems, it is applied research. Using the Mattersich definitions, it can be seen that *AJER* has interpreted its mission to publish research in education broadly enough to include studies that lie within each of the two categories of intentionality. It can also be seen that the greater number of studies lies in the application of knowledge category.

A selection of 25 of the studies on reading reported from 1955 to the present illustrates this weighting. The earlier studies such as the one by Carmichael and Rees (1955) surveyed reading achievement levels attained by Alberta pupils in order to discover whether or not they were meeting standards attained elsewhere. The results of comparisons like these were intended for use by practitioners who, presumably, would take them into account when planning curriculum and instruction in reading. An experimental study of Koziey and Brauer (1972) described methods of teaching reading which demonstrably improved reading performance. These and other studies in this particular group of articles were clearly intended to assist practitioners with the resolution of problems encountered by teachers in the teaching of reading or by pupils in learning how to

read. By comparison there are relatively few articles reporting research deliberately devised to discover new knowledge about reading and language phenomena. Stewin and Anderson (1974) tested the power of the concept of cognitive complexity in information processing; Evanechko, Armstrong and McFetridge (1974) explored the semantic organization underlying verbal behaviour; Cummins and Gulutsan (1974) focused on cognitive growth; and Smith (1973) searched for understanding in terms of language acquisition concepts.

Whether *AJER* articles report findings from applied or from basic research studies, one is struck by the consistent emphasis placed on matters of method and technique in the majority of them. Perhaps this is not too surprising in view of the metamorphosis which has taken place in social science research in the post-war years. The earlier volumes of *AJER* contain reports of studies that are predominantly descriptive presentations. These are succeeded by reports of studies which used analytical-empirical approaches in the attempt to adapt scientific methods to education problem-solving. From the mid-1960's on, there is concern for more rigour, and for more emphasis on the development of theories and models (Holdaway & MacIver, 1966), methodology (Maguire, 1969), and new techniques (Hunka, 1971). It was during this period, according to Gephart (1980), that researchers in education became preoccupied with five major types of problems in the conduct of research: problems of analysis, treatment, representativeness, measurement, and logic, which even yet have not been satisfactorily resolved. One detects in a few of the articles published after 1970 the adoption of a different philosophy of science, one which takes into account the researcher as person, an integral part of the research enterprise. Nevertheless, the research methods and techniques employed most frequently in the articles appearing in *AJER* are those which were developed originally for studies in agriculture, biology, or psychology.[1] It is exceptional to find reports of studies making use of interview, historical, case study, content analysis, opinion poll, or survey sampling. Of fifty-four articles published in *AJER* from 1955-1979 which designate teachers or teaching as a research focus, only six reported the use of such techniques as economic prediction, field study, Delphi study, case study, or interview procedures.

Educational research has come to be almost exclusively associated in the minds of some people with the use of systematic descriptions and measurement, observation, and controlled experimentation, or the creation of instruments to test knowledge through verification, confirmation, or falsification. Proportionately larger numbers of educational researchers in North America have been committed to the use of this particular methodology, resulting, one would expect, in a large pool of articles from which to select those for publication. Thus, while the articles in *AJER* represent the prevailing trends in educational research for the past twenty-five years, at the same time they provide a silent commentary on the comparatively limited number of research studies conducted using other forms of inquiry. This makes for an imbalance that skews the image of what counts for research in education. Eastwood's (1975) comment appearing in Volume XXI seems particularly pertinent in this regard:

> No one would wish to suggest or be accused of suggesting that a technological research pattern is an adequate or appropriate model for educational research.... Human learning and human development are considerably more complex processes

> than the development of a new machine. But both are practical activities dependent upon the application and utilization of basic research and both necessitate research in ways of applying. (p. 74)

The effects of publishing educational research

The experience of delving into twenty-five volumes of *AJER* has forcibly reminded me of the diversity of questions confronting researchers in education. Some are conceptual, others are methodological, and still others have to do with the relationship between research and practice. Clearly, the raising of such questions as a stimulus for further thought and research activity is one of the primary functions of the *Journal* and one it has performed well for its readers over the years. It is consistent with the statement of purpose set out in the masthead: *AJER* is intended to be

> A quarterly journal devoted to the dissemination, criticism, interpretation and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

My overview has indicated a few of the ways *AJER* has so admirably fulfilled these goals. It has also served to whet my appetite for more information about what the field of educational research was like at the time each of the twenty-five volumes was in process of preparation for publication. Editors of major research journals such as *AJER* read the assessments made by knowledgeable reviewers of hundreds of manuscripts; they make informed decisions about selections; they interact with authors representing research in disciplines using different forms of inquiry. I think it would be valuable to know more about the issues in educational research that consumed the interests of authors and editors of the day. I can only speculate a little about the effect that the inclusion of editorial statements might have had. If editorials serve to highlight issues important to the mission of a given publication, might the editors of *AJER* have been able to communicate some sense of the issues and events and the world of ideas in educational research? Might they have sharpened insights or created linkages that could have intensified or multiplied the effects of ongoing research? Would such editorials have helped us to appreciate their perceptions, so well described by Hodysh (1971) as the compound of needs, values, hopes, past experience, and culture used to organize and to direct interpretations of reality?

I applaud the invitation first extended in Volume 24 to provide opportunities for scholars to write critically about the general conduct of research in education (Hodysh, 1978). Essay reviews, book reviews, and expressions of points of view in "Perspectives" pieces are being included in a deliberate effort to present the questions being raised in the larger world of research, the debates about premises, the shifts in what is taken for granted in educational research. Even so, there are problems in publishing matters of current concern; it takes time for the cycle of the research process to come round to the publication stage. This publication time-lag continues be a dilemma in disseminating ideas and research results because of the benefits that accrue from criticism and revision prior to publication. The provision for alternative types of articles should encourage those who would challenge our existing ways of thinking and those who can generate excitement about the search for knowledge.

But having said these things, I must conclude by acknowledging the far-reaching positive effects of the research articles published by *AJER*. There is very substantial evidence in articles appearing over successive years demonstrating that the conduct of research in education has been facilitated by publication of work leading to the refinement and improvement of measures and methods. There is also evidence along the lines Suppes (1978) has identified for North America generally, that knowledge derived from basic research has had an effect on practice in education: policy decisions about the content and organization of curricula in science and mathematics have been influenced by Piaget's notions of conservation; certain expectations about students and teachers have been modified in light of findings in social psychology; and work in the social sciences and humanities has changed some concepts about the training of administrators, to take just a few examples. *AJER* came into being at a time when the vision for the development of educational research in Canada was fresh and exciting. It is impossible to assess precisely the extent to which the *Journal* helped to translate that vision into reality, or to gauge exactly the extent to which the vision itself shaped the development of *AJER*. Traces of both of these factors are to be found throughout the twenty-five volumes of the *Journal*. It is a singularly fine record of many of the achievements and of some of the struggles in educational research in Canada and abroad.

And what of the next twenty-five years for *AJER*? At a time when large numbers of publications in every field are in peril of disappearing, my hope is that *AJER* will continue to flourish. From its earliest beginnings to the present time, it has played an important role in the whole of the educational research enterprise in Canada. It may be that this role can be sharpened somewhat. The time seems propitious, with the appointment of the panel of widely representative Consulting Editors, for the *Journal* to assume responsibility for actively identifying gaps in knowledge and needed research in education. *AJER* might deliberately assume responsibility for bringing together a critical mass of knowledge from which the members of the teaching profession as well as scholars in education could draw applications. Or, perhaps, the *Journal* might consciously assume a special responsibility for disseminating the results of research studies that will further, in the broadest sense, our understanding of teaching and schooling and the education of practitioners.

The resilience demonstrated by *AJER* in its first twenty-five years of publication will surely stand it in good stead now. It has been characterized by a sense of purpose and by those qualities of vitality and perseverance which are needed to sustain it through periods of change. These are qualities that will enable *AJER* to continue to make important contributions to the study and practice of education in Canada in the years ahead.

Notes

1. For a useful comparison, see V. L. Willson, Research Techniques in *AERJ* articles: 1969-1978. *Educational Researcher*, 1980, 9(6), 5-10.

References

Aikenhead, J. D. To teach or not to teach. *AJER*, 1956, *2*(3), 173-185.

Bana, J. P., & Nelson, D. Some effects of distraction in nonverbal mathematical problems. *AJER*, 1977, *23*(4), 268-279.

Baxter, H. University research in jeopardy: The threat of declining enrolment. *CAUT Bulletin*, 1980, *27*(3), 3-4.

Bourgeois, R., & Nelson, D. Young children's behavior in solving division problems. *AJER*, 1977, *23*(3), 178-185.

Carmichael, A., & Rees, R. E. A survey of reading achievement in Alberta schools. *AJER*, 1955, *1*(1), 18-33.

Cheong, G. C., & Wadden, E. The relationships between teachers' experimental and dogmatic attitudes and their pupils' self-concept. *AJER*, 1978, *24*(2), 121-125.

Clarke, S. C. T., & Dunlop, G. M. Tests in measurement and evaluation. *AJER*, 1956, *2*(1), 57-59.

Clarke, S. C. T., & Pilkington, W. Why teaching is chosen as a career. *AJER*, 1955, *1*(1), 44-52.

Côté, A. D. J. Piaget's recent formulations on perception and intelligence. *AJER*, 1967, *13*(4), 299-310.

Cowper, D. J., & Stewin, L. An examination of the illusion phenomenon in Soviet set theory: An experimental critique. *AJER*, 1974, *20*(1), 39-45.

Crawford, D. H. Piaget and school mathematics. *AJER*, 1960, *6*(3), 125-136.

Cropley, A. J. Creativity, intelligence, and achievement. *AJER*, 1967, *13*(1), 51-58.

Cummins, J., & Gulutsan, M. Bilingual education and cognition. *AJER*, 1974, *20*(3), 259-269.

Eastwood, G. R. Educational research: Concepts and models. *AJER* , 1975, *21*(2), 71-83.

Eddy, W. P. Teacher characteristics and Social Studies achievement. *AJER*, 1963, *9*(4), 254-263.

Evanechko, P. O., Armstrong, R. D., & McFetridge, P. A. Semantic space and the development of word meaning. *AJER*, 1974, *20*(4), 305-315.

Frost, B. P. Intelligence, manifest anxiety and scholastic achievement. *AJER*, 1965, *10*(3), 167-175.

Gephart, W. J. Research: A five-faceted problem-solving process. *CEDR Quarterly*, 1980, *13*(1), 23.

Greenfield, T. B. Research on the behavior of educational leaders: Critique of a tradition. *AJER*, 1968, *14*(1), 55-76.

Greenfield, T. B., & Andrews, J. H. Teacher leader behavior as related to pupil growth. *AJER*, 1961, *3*(2), 93-102.

Guba, E. G. *Toward a methodology of naturalistic inquiry in educational evaluation.* Los Angeles: Centre for the Study of Evaluation, UCLA, 1978.

Gulutsan, M. Jean Piaget in Soviet psychology. *AJER*, 1967, *13*(3), 239-248.

Hobbs, E. D. Methodological problems in conservation testing with particular reference to volume conservation. *AJER*, 1975, *21*(4), 262-277.

Hodysh, H. W. Cognitive perception and the conceptual foundations of research in educational history. *AJER*, 1971, *27*(2), 113-116.

Hodysh, H. W. Call for papers. *AJER*, 1978, *24*(4), 275.

Holdaway, E. A., & MacIver, D. A. An analysis of the use of models in education. *AJER*, 1966, *12*(3), 163-188.

Hoxter, L. Teacher attitudes and cultural differences. *AJER*, 1974, *20*(2), 133-145.

Hritzuk, J., & Janzen, H. A Soviet and Western concept of set. *AJER*, 1970, *16*(4), 217-225.

Hunka, S. M. Instructional applications of the computer: An experimental course. *AJER*, 1971, *17*(1), 55-62.

Keeler, B. T., & Andrews, J. M. Leader behavior of principals, staff morale and productivity. *AJER*, 1963, *9*(3), 179-194.

Kieren, T. E., & Nelson, D. The operator construct of rational numbers in childhood and adolescence. *AJER*, 1978, *24*(1), 22-30.

Koziey, P. W., & Brauer, J. H. Using mental practice to improve reading performance. *AJER*, 1972, *17*(3), 190-195.

Lavoie, J. C., & Adams, G. R. Teacher expectancy and its relation to physical and interpersonal characteristics of the child. *AJER*, 1974, *20*(2), 122-132.

Lefrançois, G. A theory of behavior. *AJER*, 1967, *13*(2), 121-128.

Lefrançois, G. Jean Piaget's developmental model: Equilibration-through-adaptation. *AJER*, 1967, *13*(3), 161-172.

Liedtke, W. W., & Nelson, L. D. Concept formation and bilingualism. *AJER*, 1968, *14*(4), 225-232.

Lindstedt, S. A. Teacher qualifications and grade IX mathematics. *AJER*, 1960, *6*(2), 76-85.

MacDonald, J. A survey of Soviet education. *AJER*, 1959, *5*(2), 100-111.

MacDonald, J. Explanation and the neo-Marxist theory of mental testing. *AJER*, 1960, *6*(1), 54-59.

Maguire, T. O. Decisions and curriculum objectives: A methodology for evaluation. *AJER*, 1969, *15*(1), 17-30.

Mattessich, R. On the essence of basic and applied research in the administrative and other instrumental sciences. *University Management, Education and Research: A developing crisis*. Ottawa: Social Sciences and Humanities Research Council of Canada Report, April, 1980. Cat. No. CR23-2/1980E.

McBride, B. E. J. A study of student assessment of teacher performance. *AJER*, 1964, *10*(2), 79-89.

McLeish, J. The Russian Sunday schools: An experiment in voluntarism. *AJER*, 1974, *20*(1), 75-95.

Miller, P. J. Student and teacher attitudes towards education: An exploratory study. *AJER*, 1974, *20*(3), 251-258.

Nelson, L. D. Textbook difficulty and mathematics achievement in junior high school. *AJER*, 1964, *10*(2), 105-113.

Nelson, L. D., & Kieren, T. E. Children's behavior in solving spatial problems. *AJER*, 1977, *23*(1), 22-30.

Nelson, L. D., & Towler, J. O. Spatial concepts of elementary school children. *AJER*, 1967, *13*(1), 43-50.

Nelson, L. D., & Worth, W. H. The mathematical competence of prospective elementary teachers. *AJER*, 1960, *6*(1), 3-9.

Noll, R. L., Willower, D. J., & Barnette, J. J. Teacher self-actualization and pupil control ideology-behavior consistency. *AJER*, 1977, *23*(1), 65-70.

O'Bryan, K. G., & MacArthur, R. S. A factor analytic study of Piagetian reversibility. *AJER*, 1967, *13*(3), 211-220.

Poole, H. E. A logically first stage in training creative teachers. *AJER*, 1975, *21*(3), 143-153.

Powell, F. A critical evaluation of the Detroit Beginning First Grade I.Q. Test. *AJER*, 1956, *2*(1), 30-40.

Radnitzky, G. *Contemporary schools of metascience*. (3rd ed.) Chicago: Henry Regnery, 1973.

Randhawa, P., Hunt, D., & Rawlyk, S. A. Factorial structure, reliability, validity, effectiveness and efficiency of the Canadian Cognitive Abilities Test. *AJER*, 1974, *20*(2), 208-215.

Safran, C. The Safran Culture Reduced Intelligence Test. *AJER*, 1963, *9*(1), 36-44.

Sangster, C. H. An evaluation of the effectiveness of a standardized test of language. *AJER*, 1956, *2*(3), 186-202.

Shulman, L. (Ed.) *Review of research in education.* Volume 4. Itasca, Illinois: F. E. Peacock, 1976.

Smith, P. R. Soviet studies in language acquisition. *AJER*, 1973, *19*(2), 109-118.

Sparby, H. T. The inter-disciplinary approach to the training of school administrators. *AJER*, 1959, *5*(2), 87-96.

Stewart, F. K. Preface. In *Educational administration in Canada.* Toronto: Macmillan, 1970, pp. *iii, iv.*

Stewin, L. L. The Soviet theory of set. *AJER*, 1969, *15*(2), 95-102.

Stewin, L. L. Integrative complexity: Structure and correlates. *AJER*, 1976, *22*(3), 226-236.

Stewin, L. L., & Anderson, C. C. Cognitive complexity as a determinant of information processing. *AJER*, 1974, *20*(3), 233-243.

Stewin, L. L., & Martin, J. The developmental stages of L. S. Vygotsky and J. Piaget: A comparison. *AJER*, 1974, *20*(4), 348-362.

Suppes, P. (Ed.) *Impact of research on education: Some case studies.* Washington, D. C.: National Association of Education, 1978.

Swift, W. H. *Educational administration in Canada.* Toronto: Macmillan, 1970.

Vernon, P. E. The validity of divergent thinking tests. *AJER*, 1972, *18*(4), 249-258.

Walker, B. E. Fifty years after Alfred Binet. *AJER*, 1959, *5*(3), 173-175.

Willson, V. L. Research techniques in *AERJ* articles: 1969-1978. *Educational Researcher*, 1980, *9*(6), 5-10.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 472-473*

Henry W. Hodysh

Educational Research in Transition: An Historical Perspective on Editorial Change

The mandate of *The Alberta Journal of Educational Research* (*AJER*) to critique and disseminate the results of educational inquiry has focused the attention of the journal's editors ever since its appearance in 1955 (Coutts, May 1993, personal interview). For Canada's first journal of educational research, this commitment engaged a broad cross section of the academic community.

From *AJER*'s inception (Buck, 1994), its editor H.S. Baker had the support of the Alberta Advisory Committee on Educational Research and the Faculty Committee on Educational Research, including such individuals as H.T. Coutts, S.C.T. Clarke, and H.E. Smith (*AJER*, 1955). A concern for the journal's intellectual and organizational well-being was indicated in G.M. Dunlop's plea for "wider assistance and cooperation of interested individuals throughout the province" (1956, p. 6). This initiative was furthered in 1965 when an Editorial Committee under G.R. Eastwood introduced a new set of editorial policies (1965), leading in 1966 to the formation of an Editorial Board comprised of professors from the University of Alberta and the University of Calgary (*AJER*, 1966). One year later, P. Lane assumed editorial responsibility for "the discussion and disposition of manuscripts" (*AJER*, 1967). She was followed in turn by editors E. Miklos, J. McLeish, T. Kieren, and A. Clark. By the early 1970s, the efforts of the editors and their associates had already contributed to a wide readership and, according to the publications assessors of the Canada Council, a well-deserved reputation as a leading journal of educational research (Canada Council, 1971).

When I became editor of *AJER* from 1977 to 1981, my twofold intention was to heighten its international profile and provide for different approaches to the interpretation of educational research and debate. To achieve the first objective, a panel of Consulting Editors was established to assist in the review of manuscripts (Hodysh, 1978a). The panel, representing scholars from Europe, the United States, Canada, Great Britain, and Australia, provided expertise in a variety of educational areas that extended beyond the valued assessments of the journal's Faculty Publications Committee.

During this period, two new categories for the publication of educational research were introduced. In a "Call for Papers," manuscripts were invited for "Perspectives" and "Essay Reviews" (Hodysh, 1978b). Presenting a "speculative or philosophic point of view," the purpose of the "Perspectives" manuscript was to challenge fundamental understandings of research by suggesting alternative directions of thought and argument. The "Essay Reviews" called for a comprehensive analysis of recently published educational research, transcending the depth of analysis already evident in *AJER*'s "Book Reviews" section.

The attempt to foster continuing debate among scholars also led to the introduction of "Research Notes" (Nyberg & Nyberg, 1980) and "Rejoinders" (Wright & Elliott, 1981; Parsons, 1981; DeFaveri, 1981). Whereas the "Research Notes" allowed for a short but focused critique of a problem or technique in research, the "Rejoinders" were limited to a brief exchange of views usually between an author of an article and a critic, reviving a tradition founded in an earlier edition of *AJER* (Anderson, 1960).

Whether these developments have advanced the cause of educational research is possibly a judgment of time. What they do suggest is the need for each generation of educators to define a journal in its own way. A journal's concern with such questions as the relative significance of basic and applied educational study and the direction research should assume becomes, however, more than a decision of any one individual. It reflects a recognition of what has gone before and, more importantly perhaps, what scholars today both in and outside of the educational community believe is of most worth.

References

AJER. (1955). Faculty Committee on Educational Research. *1*(1), inside cover.
AJER. (1966). Editorial Board. *12*(4), inside cover.
AJER. (1967). Editorial Board. *13*(4), inside cover.
Anderson, C.C. (1960). Dr. Macdonald and mental testing: A rejoinder. *AJER, 6*(1), 60-63.
Buck, G.H. (1994). Herbert T. Coutts and the origins, early development, and possible future directions of the *Alberta Journal of Educational Research*. *AJER, 40*(1), 3-6.
Canada Council. (1971). Letter to editor of *AJER*. Edmonton: Organizational Files, Publication Services, Faculty of Education, University of Alberta.
DeFaveri, I. (1981). Rejoinders. Reply from DeFaveri. *AJER, 17*(1), 101-102.
Dunlop, G.M. (1956). Editorial. *AJER, 2*(1), 5-6.
Eastwood, G.R. (1965). Editorial. *AJER, 11*(4), 201.
Hodysh, H.W. (1978a). The Alberta Journal of Educational Research and Consulting Editors. *AJER, 24*(4), 221-222.
Hodysh, H.W. (1978b). Call for papers. *AJER, 24*(4), 275.
Nyberg, V.R., & Nyberg, A.M. (1980). Research notes. Reliability of the Alberta essay scales. *AJER, 26*(1), 64-67.
Parsons, J. (1981). Rejoinders. A reaction to "Moral education: The risk of oversimplification." *AJER, 27*(1), 99-100.
Wright, I., & Elliott, M. (1981). Rejoinders. Moral education: A reply to DeFaveri. *AJER, 27*(1), 95-98.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 474-474*

Walter H. Worth
Editor, March 1985-December 1988

Perceptions

There are three reasons for advocating the reprinting of Neil Johnson's (1987) article that appeared in the "Perspectives" section of *AJER* a decade ago.

One is to highlight the desirability of *AJER* continuing to provide a forum for the publication of articles that offer a point of view or state-of-the-art review about an issue or concern in educational research. Such reflective or interpretive essays can help in overcoming conceptual and methodological weaknesses in conventional research studies.

Another is because the article focuses on the basic ingredient of educational research—perceptions. These preferred approximations of reality as defined by researchers and their subjects shape both the processes and the outcomes of our endeavors. Accordingly, it is imperative that we understand what perceptions are, and how they develop and influence behavior. Equally imperative is our need to use such knowledge to devise methodologies that enhance the trustworthiness of scholarly inquiry.

Johnson's article, written when naturalistic research was becoming trendy, addresses these twin imperatives. It examines some current developments in perception theory and research as a basis for generating some generalizations about perceptions. The pertinence of these generalization for both consumers and producers of research is then discussed, including an examination of means to limit the potentially prejudicial impact of perceptions on educational research.

A further reason for drawing attention to this article is to encourage someone to prepare an update on the current state of our knowledge about perceptions and how to cope with them in ways that will enhance the trustworthiness of our investigations.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 475-497*

PERSPECTIVES

The Pervasive, Persuasive Power of Perceptions

Neil A. Johnson
University of Alberta

Through their influence on attitudes and behavior, perceptions are central to both practice and research in education. This article examines some leading developments in perception theory and research, particularly in respect of (a) the influence of perceptions on behavior, (b) the accuracy of perceptions, (c) the capacity of individuals to express perceptions, (d) perception processes, and (e) major factors influencing perceptions of persons and social events. This discussion leads to the presentation of seven generalizations about perception and a consideration of their implications for all who engage in educational endeavors. Most importantly, it serves as a basis for examining means to limit the potentially prejudicial impact of perceptions on educational research.

Perceptions shape human attitudes and behavior; their impact is pervasive and unavoidable. They provide bases for understanding reality—objects, events, and the people with whom we interact—and our responses to them. Thus perceptions dominate all the situations that educational and other social researchers study. Perceptions also focus the work of the researchers themselves—the behaviors researchers notice and choose to record, the opinions they invite from others, and the theoretical perspectives within which they interpret data and frame conclusions.

An initial intent of this article is to draw attention to extant knowledge about the nature of perception and the factors that affect the perceptions of individuals in educational organizations. This provides a basis for advancing a framework of factors associated with perception and for considering important implications of perception theory for educational researchers and practitioners. Researchers, in particular, need to be conscious of the pitfalls inherent in relying on their own and others' perceptions.

Reprinted from *The Alberta Journal of Educational Research*, 33(3), 206-228, 1985.

The ideas presented were drawn from the literature, from the author's discussions with researchers and educators, and from personal experience.

The Pervasive Influence of Perceptions on Educational Practice and Research

Practitioners in educational settings deal with perceptions constantly. They are first employed and then evaluated substantially in accordance with perceptual impressions. Instructional practices, administrative decisions, and policy initiatives are guided by perceptions of individuals and educational needs, as are responses to the actions of coworkers and others. Knowledge of employees' and other stakeholders' perceptions helps administrators to revise educational policy and change individuals' experiences in educational organizations, and knowledge of perception theory may provide an avenue for directly improving educational leadership and practice.

Educational research also involves scholars in these situations. Seeking knowledge of the interactive behavior, personalities, attitudes, and emotions of practitioners, researchers enter university and school classrooms, the offices of administrators, and a variety of other locations to elicit perceptions from policy makers, administrators, instructors, and students. For those investigators who desire no more than understanding of attitudes and feelings, perceptions are paramount. And, because behavior occurs in response to perceptions, perceptual data also hold a key to researchers' knowledge of organizational behavior. Some scholars use questionnaires and interviews to gather perceptual data for quantitative analysis. Naturalistic researchers and policy analysts also rely on perceptions stated by others, and they document their own perceptions of organizational events and environments. Researchers using interpretive orientations value perceptions as indicators of the realities for individual educators. Conceding that an objective reality can be known only through the filter of perceptions, researchers of a rationalistic bent also depend on perceptions—as surrogate measures of reality. Doubting the trustworthiness of their data, some scholars explore perceptions still further, using a variety of sources, strategies, and investigators to corroborate their research findings as true and full representations of life in educational settings.

If educational research is to make important contributions to knowledge, researchers need to become conversant with the nature and processes of perception and with the factors that shape and can bias the perceptions of those engaged in research and practice. This will allow them to address personally and methodologically the trustworthiness of their scholarly endeavors.

The Nature of Perception

A Problem with Definition

To date, the notion of perception has resisted clear and conclusive definition. Once regarded simply as "sensation plus meaning" (Bartley, 1980, p. 5), most psychologists now view sensation as a physiological experience and perception as a cognitive activity (Krech, Crutchfield, Livson, Wilson, & Parducci, 1982). Perception and other cognitive activities such as "inferring, categorizing or judging" (Tajfel, 1969, p. 316). Despite this conceptual uncertainty, there is broad consensus that perception is "the understanding of the world that you construct from data obtained through your senses" (Shaver, 1981, p. 83). As Shaver noted, such a definition implies that perceptions are obtained through sensory experiences rather than merely by reflection or intuition, that an objective world exists outside the perceiver, and that the perceiver actively forms an impression from each stimulus.

Developments in Theory

Investigation of physiological sensory processes (e.g., Bartley, 1972; Hochberg, 1978) has been unable to account fully for perceptual experiences, so many psychologists have paid closer attention to the complementary psychological dimension. To this end, structuralists tried to expose and describe the characteristics of and relationships between aspects of conscious experience (Dember & Warm, 1979). Gestalt theorists criticized those attempts to decompose total perceptual structures and focused instead on how individuals perceive situations as a whole (Moates & Schumacher, 1980). Allport (1955, pp. 576-612) extracted from 13 "specialized, divergent, and often discordant" theories of his time eight major generalizations (most of which survive in current theories): (a) individuals aggregate and interrelate multiple understandings of events; (b) these "perceptual aggregates" are organized within limiting conceptual boundaries; (c) perceptions are assembled over time; (d) although general order and stability prevail, some perceptual inconsistencies are tolerated; (e) perceptions remain relatively constant over time; (f) there is a tendency to return to original "steady state" perceptions following the disruption of new impressions; (g) impressions are weighed unequally in perceptual aggregation; and (h) although perceptual aggregates sometimes conflict, usually they mutually support higher-order perceptual generalizations. Kelly's (1955) "role-construct theory" emphasized interrelationships among the individual's perceptions, interpretations, and behaviors: By developing "constructs," or categories of events, the perceiver forms expectations and plans behavioral responses; and comparison of new perceptions with existing systems of constructs validates current understandings and classifications of people, objects, and events. Likening perceptual processes with computer system procedures, however, "information-processing" theorists contended that perception occurs only after sequential transformation stages: stimulus, sensory store, perceptual image, memory, and response. Moates and Schumacher (1980, p. 276) noted that, whereas Gestalt psychologists "had a remarkable flair for capturing the intuitive aspects," information-processing theory helped to "formalize the sometimes vague gestalt ideas so that they are more testable." The study of "social perception," as yet still "in its infancy," focuses specifically on three important avenues of research: (a) study of the effects of social interaction on perceptions of physical properties; (b) investigation of perceptions of individuals' physical and social characteristics; and (c) exploration of perceptions of social relationships and events (Hochberg, 1978). The latter, of primary interest to educational researchers, includes the emerging fields of attribution theory and implicit personality theory.

The Perception Process

Perceptions seem to be formed in a series of cognitive steps. A common view is that a perceiver selects and categorizes sense data within predetermined structures, or frames of reference; in turn, these are subject to attributes of personality (Kelly, 1980).[1] According to Bruner's (1951) "expectancy or hypothesis theory of perception," three stages are involved: the individual hypothesizes about the occurrence of a likely event ("hypothesis"), the environment provides an informational stimulus ("information"), and this prompts a confirmatory response from the individual ("confirmation") (pp. 121-127). Moates and Schumacher (1980, pp. 16-17) advanced a more detailed explanation: sensory receptors are oriented toward a source of stimulation; certain features and contextual factors are extracted (noticed); and then the perceiver engages in "a cyclic process of orientation, feature extraction, comparison with memory, and then additional orientation, feature extraction, and comparison," permitting

the perception to be incrementally refined. This "chain of perception" process (Krech et al., 1982, p. 103) accords with Litterer's (1973, p. 108) "selection-closure-interpretation" explanation and Forgus and Melamed's (1976) description of a sensation-perception cycle.

Four questions about this perception process are of central concern to educators and researchers: (a) To what extent do perceptions affect behavior in educational and other social settings? (b) Are members of educational organizations able to consciously identify and express their perceptions? (c) How accurately do perceptions portray reality? and (d) Are perceptions shaped by identifiable and commonly occurring factors?

Perceptions and Behavior

Through a process of "discovering what the environment is really like and adapting to it" (Neisser, 1976, p. 9), perceptions shape the social behavior of individuals. Perceptions allow individuals to understand, anticipate, and react to environmental circumstances, events, and the behavior of others (Forgus & Melamed, 1976; French, Kast, & Rosenzweig, 1985; Harvey, Weary, & Stanley, 1985; Wrightsman, 1977). Blake and Ramsey (1951), Litterer (1973), and Kelly (1980) also highlighted perceptions as the critical determinants of behavior in organizational settings and, although Tagiuri (1969, p. 435) warned that other factors also impose on social behavior, he concluded that, "if there is to be a science of interpersonal behavior, it will be based, to some extent, on our learning more about how people come to perceive other people as they do." Hochberg (1978, p. 242) also regarded the study of perception as "an important tool for understanding and predicting behavior" in social situations, and Shapiro and McPherson (1987, p. 75) recently focused attention on public policy makers' perceptions of policy dilemmas as "an important determinant" of their "policy behavior."

Understanding of perception is therefore critical for educational research and practice, for social behavior in educational settings is guided not merely by an assumed objective reality but by actors' individual perceptions and by the factors that shape and distort those perceptions.

Capacity to Express Perceptions

Though instrumental in determining behavior, many perceptions are beyond the capacity of individuals to consciously recognize and verbally express. Cameron and Whetten (1983, pp. 12-13), for example, concluded from organizational effectiveness research that

> There appears to be ample empirical evidence . . . to suggest that individuals frequently cannot report accurately the criteria of organizational effectiveness that they implicitly hold. Nor are they aware of the factors that motivate their judgments or evaluations of an organization. When researchers ask various constituency members to specify important criteria of effectiveness, there is no assurance that the criteria they enumerate will be consistent with the criteria they use implicitly to judge effectiveness.

Cameron and Whetten supported this view by citing research that asked organization members to rank "47 different goals in their institutions"; none received particular priority. Clearly, these authors' assessment was based on an especially onerous demand on the recall and cognitive organizational capacities of respondents. Nevertheless, it does signal a potential danger in seeking perceptual measures of individuals' impressions about abstract organizational attributes.

How Accurate are Perceptions?

On the matter of perceptual accuracy, Rock's (1975, p. 5) research findings are encouraging: "granting that the perceived world is different from the world that is the object of perception, one can still say that there is a high degree of correspondence." If so, we may treat perceptions as broadly accurate indications of the "real worlds" of educational organizations. On the other hand, the literature also includes numerous references to "illusions" attesting to a clear discrepancy between reality and perceptions of objects. Of course, the accuracy of perception of social events is more difficult to measure, for phenomena that defy direct inspection and measurement can be known and understood only through a process of perceiving; if individuals' perceptions of an event agree, we can only assume that they reflect reality (Rock, 1975; Shaver, 1975). Even so, Gladstein's (1984) research and a range of studies reviewed by Wrightsman (1977) disclosed a pervasive human incapacity for perceiving and recalling either objective or social events with accuracy—a finding that supports Allport's (1961) generalization that "good judges" are rare and Blake, Ramsey, and Moran's (1951, p. 5) conclusion that "sometimes . . . it is not difficult to show that the margin of interpretative error is very wide." Clement (1978) blamed selective perception for this apparent disparity—selections that, of course, differ among individuals. Availability of information is an associated problem: possessing only fragmentary information, individuals are frequently obliged to make perceptual assumptions and hold personal expectations which lead them to perceptions that diverge from reality and from those of other witnesses (Hochberg, 1978; Litterer, 1973). Hochberg (1978) further identified three common causes of perception-reality discrepancy: (a) events that cannot be discerned; (b) omissions, additions, and distortions arising out of human perceptual processes; and (c) events whose significance is misunderstood. Differing histories of perceptual learning, attention, and intentions also lead individuals to form different impressions of events and persons.

As much of the research and our personal experience tell us, individuals' perceptions of social events frequently differ. For example, we may expect an instructor, an inquiry-oriented student, and a student motivated only by threat of examination to perceive the purpose, product, and quality of a shared seminar experience quite differently. Hence educational practitioners and researchers would do well to draw on a variety of perspectives and to recognize the limitations of their own and others' perceptions when forming impressions and making decisions. For many researchers, in particular, reliance on multiple data sources may increase the convergence of perceptions and thus the likelihood of discovering a shared or objective reality.

Despite inaccuracy, there is often considerable consistency in social perceptions as well as in the biases that cause individuals' perceptions to depart from reality and the impressions of others. Perception research has disclosed a range of central influences—summarized in Figure 1 and discussed below—that can help us to understand and account for the effects of perceptions on educational research and practice.

Factors that Affect Perceptions

Perceivers arrive at their limited, differing, sometimes distorted impressions of persons, objects, and events only after influence from a range of complex and subtle factors. Despite an initially bewildering array of influential factors, certain "perceptual determinants" emerge as common among perceivers (Blake et al., 1951, p. 8). Scholars have classified them in a variety of ways, such as familiarity, strength, and salience of stimuli (Kelly, 1980) and "experience, intelligence, and empathic ability" of perceivers (Shaver, 1975, p. 22). Most comprehensive are Forgus and Melamed's (1976) four

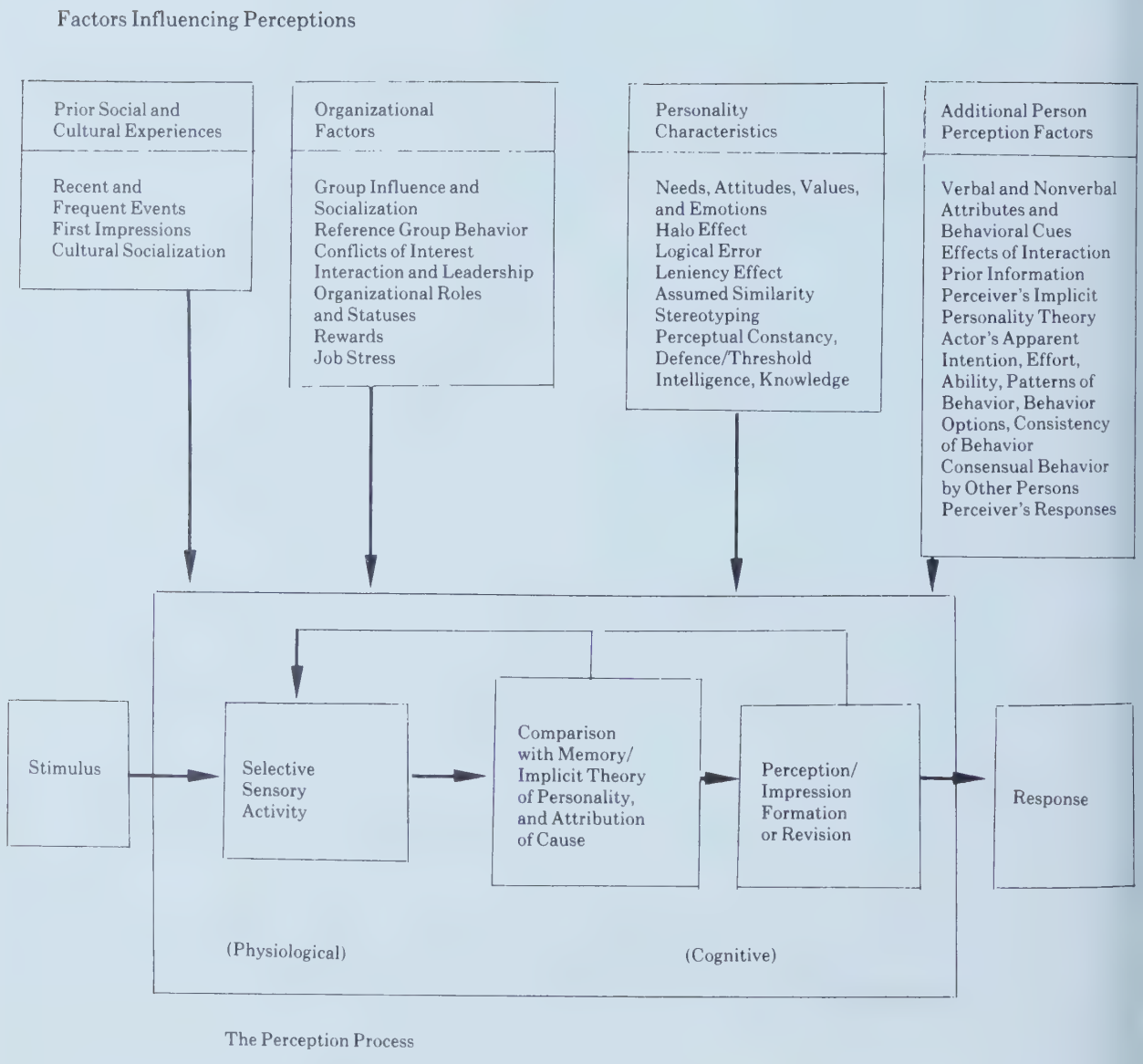


Figure 1: The Perception Process and Influential Factors

categories: (a) influences of social experience and cultural background; (b) impact of the perceiver's values, attitudes, and personality; (c) dynamics of person perception; and (d) perceptions of causality in social events.

Social and Cultural Experience

Social/organizational experience and cultural heritage are widely accepted as important determinants of perceptions. Clement (1978, p. 51), for example, argued that individuals process and categorize environmental occurrences according to existing "perceptual structures"—structures based on prior experience that are expanded and altered by the new perceptions. To Wrightsman (1977, p. 95), these perceptions develop as if "peeling an onion." Though the weighting of new and existing pieces of information is unclear, first impressions and recent and frequent events seem to impose heavily on subsequent impressions—except where perceivers consciously resist their prejudicial influence (Epstein & Rock, 1960; Krech et al., 1982; Litterer, 1973). Common work and other experiences also lead individuals to form similar perceptions (Litterer, 1973) and there is evidence that social contact and familiarity with different classes of persons and situations, group norms, personal self-confidence, and linguistic classificatory labels also affect perceptions (Tajfel, 1969).

Furthermore, perceptions of facial and other perceptual cues seem to be culturally influenced; individuals from different cultures sometimes express emotion differently, so their cues are open to misinterpretation (Brunswik, 1944; Tagiuri, 1969). Of course, the discipline of anthropology also has an extensive literature attesting to the pervasive and powerful effect of culture on individual perceptions and on resultant "appropriate" behavior. According to Murphy (1979, pp. 14, 22-23), the "majority view among social anthropologists" favors substantial cultural determination of individual perceptions and actions. By shaping perceptions, culture "tells us how we should act . . . [and] what we can expect of the other person. Behavior is thus rendered predictable, often within broad limits, and we gain a degree of mastery and confidence in social situations." Likewise, Vivelo (1978, pp. 16-17) saw culture as a "system of rules or a pattern for behavior—a conceptual pattern that individuals use to explain and justify their experiences or events." Again, Geertz (1979, pp. 43, 46-47) proposed culture as "a set of control mechanisms—plans, recipes, rules, instructions" that impose on individual perspectives and behavior. At the same time, Geertz focused attention on culture's diverse effects on individual perceptions and behavior: "Becoming human is becoming individual, and we become individual under the guidance of cultural patterns."

Organizational Factors

Litterer (1973) highlighted certain characteristics of organizational life that affect members' perceptions. Particularly important is organizational socialization: colleagues, superordinates, and subordinates instruct perceivers about physical, financial, personal, and other signals, thereby educating them about prevailing views of reality and encouraging them to behave accordingly. Individuals—particularly those who lack self-confidence or who value social acceptance—are persuaded to accept, or at least appear to accept, the "group consensus," especially in ambiguous social settings. Perceptions are further affected by the nature of organizational interaction, by authoritative or participative leadership, by expected and actual role behavior and statuses, and by rewards for creativity and achievement; and work stress can prompt unduly early and inaccurate impression formation. The "reference groups" individuals use to judge norms of acceptable behavior and to provide bases for comparison

also influence perceptions and behavioral responses. As French et al. (1985) added, groups frequently distort perceptions about opposing groups with competing interests, leading to mutual mistrust, suspicion, power-consciousness, competitive behavior, and misunderstanding of others' intentions.

Personality Characteristics

Many personality variables have emerged as important influences on individuals' social perceptions; beliefs, needs, emotions, attitudes, values, and expectations of perceivers are frequently cited (French et al., 1985; Kelly, 1980; Krech et al., 1982). Central, however, is Tagiuri's (1969) classification of the personality-related "judgmental errors" (Krech et al., 1982, p. 705) identified by Bruner (1951) and other researchers during the 1950s and 1960s: (a) "halo effect," or a tendency to form global impressions based on overall judgments of goodness or badness (such as where a teacher obtains an initially unfavorable impression of a colleague or student and attributes other undesirable characteristics to that person); (b) "logical error," or assumed coexistence of particular traits (where a pupil assumes, say, that a teacher's energy will be accompanied by aggressiveness); (c) "leniency effect," or a tendency to rate persons high on favorable attributes and low on unfavorable ones; and (d) "assumed similarity," where other persons are assumed to have emotions and traits akin to those of the perceiver. Modern writers on perception commonly add a fifth factor: "stereotyping," that is, "a tendency to assign to people many of the characteristics (real and imagined) of the typical member of the group or class to which they are thought to belong" (Krech et al., 1982, p. 705). For this reason, a professor might be classified as absent-minded or a school principal regarded as an expert teacher merely on account of their status.

Other research has explored the intervening nature of personality through four related factors: "perceptual constancy," "perceptual defence," "recognition threshold," and "preperceptual selection." Individuals' stable and consistent perceptual frameworks are shaped and reinforced through heightened awareness and enhanced valuation of prominent and personally relevant events (Bartley, 1980; Kelly, 1974). To help preserve perceptual "balance," perceivers also "intellectualize" about discordant events, shaping them to fit existing understandings (Heider, 1980, p. 10; Litterer, 1973). Indeed, perceptual defences may cause perceivers not merely to amend interpretations of items but to defensively repress contradictory information from conscious recognition. Paradoxically, selection necessitates prior recognition (preperception); as Hochberg (1978, p. 216) explained, "the self we are aware of . . . sees only what the unconscious observer permits him to see." Less absolute than perceptual defences are recognition "thresholds." Perceivers readily identify important and consistent (low threshold) stimuli; incongruent items can also be recognized, but only after blatant or extended exposure.[2]

Additional personality factors intrude in the formation of perceptions. "Field-independent" persons retain their preconceived mental sets even in the face of new and conflicting evidence: new events all too easily convince "field-dependent" individuals to amend their perceptions (Hamachek, 1970). High intelligence (Kelly, 1974), "cognitive complexity" and maturity (Bieri et al., 1966; Wrightsman, 1977), inferential ability and knowledge of persons or situations (Gage & Cronbach, 1955), emotional adjustment, interests, and fields of specialization and social interaction (Taft, 1955) also affect judgments of people and social events, as do the personalities of stimulus persons, the nature of the judgmental tasks involved (Tagiuri, 1969) and the judgments and behaviors of relevant social groups (Asch, 1960; Deutsch & Gerard, 1960).

Perceptions of Persons

Perceptions of persons and their relationships and behavior are more complex—and often less accurate—than those of inanimate objects and other social events; for person perceptions embody impressions of "intentions, attitudes, emotions, ideas, abilities, purposes, traits, thoughts, perceptions, memories—events that are *inside* the person and strictly psychological" (Tagiuri, 1969, p. 396). Three aspects of "person perception" are important: (a) perceptions of observable behavior and attributes; (b) assignment of attributions to explain behavior; and (c) use of "implicit personality theories" to form impressions.

Perceptions of Observed Behavior and Attributes

Individuals utilize a range of cues and information sources to understand others' feelings, intentions, and relationships; extrinsic attributes of perceived individuals and the personality types, traits, prejudices, and stereotypes of perceivers themselves are powerful forces (Hochberg, 1978). Facial and physical characteristics of perceived persons provide particularly strong and consistent—but inaccurate—signals about personalities and emotions (Tagiuri, 1969); facial expressions seem to be interpreted similarly even across cultures (Ekman, Sorenson, & Friesen, 1969; Izard, 1971). Posture and position provide cues about both positive and negative attitudes, and averted gaze and abnormal mannerisms such as hesitation and silence create negative impressions of persons (Mehrabian, 1972). The implications of these findings for perceptions of self-conscious pupils, socially inept administrators, and unconventionally attired instructors are clear. Tagiuri (1969) noted that perceptions of persons are further complicated by the influences of the halo effect, logical error, and other factors referred to earlier, by role relationships, status differences, and other contextual variables, and by the differing perceptual skills needed to gain impressions of individuals, groups, and social events. Through "dyadic interplay," perceivers' own actions also affect the behavior, and thus perceptions, of those they seek to appraise:

> The perceiver . . . has to allow for the fact that he himself . . . is also the object of perception and thought and that, as such, he influences his own object of perception. Observer and observed are simultaneously observed and observer. Their reciprocal feedback processes modify their self-presentation and, in turn, their reciprocal perceptions, in a continuous recycling but varying process during which each person uses the variations in himself and the other person as a means of validating his hypotheses about the other. (Tagiuri, 1969, p. 426)

This quotation reminds us also of the array of "interaction ritual" and "impression management" strategies that we employ daily to influence the perceptions of others in the university, school, or office (Goffman, 1967; Schlenker, 1980).

To deal with all of this complexity, perceivers stereotype groups of persons and social settings; perceptions of faculty members, say, tend to be affected by impressions of outstanding professors—although observers' confidence in those judgments diminishes with the discovery of intragroup variation. More generally, four "experiential principles" seem to guide most perceptions of behavior and attributes: (a) initial person perceptions are "gut-level," generally positive or negative feelings that usually are subsequently confirmed; (b) perceivers initially notice and seek explanations for unusual attributes or behaviors; (c) first impressions focus on observed characteristics and behaviors, although perceivers quickly progress to perceptions based on personality traits; and (d) individuals view other persons in terms of generally "unified, organized collection[s] of traits that usually 'hang together'" (Krech et al., 1982, pp. 699-700).

Rist's (1970) classic study of teachers' expectations and student achievement graphically illustrates the profound educational impact of perceptual influences such as the halo effect and first impressions. Rist observed that a kindergarten teacher's classification of pupils as "fast learners" and "slow learners" was based on five initial "subjectively interpreted attributes and characteristics" of students—physical appearance, body odor, skin color, forthrightness of behavior, and use of language in classroom interaction. Having classified the pupils according to her social class-based "ideal type" of the successful student, the teacher (and soon also the students) treated the two groups differentially—encouraging one and controlling the other—and their performances diverged accordingly. At succeeding grade levels, use of these performance measures helped to reinforce the self-fulfilling prophecy that had arisen out of those initial perceptions.

Attributing Cause to Explain Behavior

According to proponents of "attribution theory," perceivers not only observe and describe persons and their behaviors, but they often seek greater understanding by attempting to determine causes for behaviors (Shaver, 1975). Attribution theory has attracted attention among social psychologists since the 1970s (Hewstone, 1983): educational researchers continue to ignore this avenue for investigating perceptions and behavior among educational policy makers, administrators, educators, and students (Frasher & Frasher, 1981). "An attribution is an inference about why an event occurred or about a person's dispositions" (Harvey & Weary, 1981, p. 6). Wrightsman's (1977, p. 100) illustration is instructive:

> People perceive the behavior of others as being caused and . . . they attribute the cause either to the person, to the environment, or to a combination of the two. For example, Joe failed to call me last night. I conclude that either he forgot (called a personal or dispositional attribution) *or* his car stalled and he couldn't get to a phone (environmental or situational attribution) *or* he was mad at me but when he got over being mad the phone at his house was tied up (combination of dispositional and situational attributions).

Apparent accidents, involuntary actions, and unintended outcomes usually require no attributions of cause, and adherence to social norms and conventions also explains many recurring behaviors. Unanticipated and seemingly intentional behaviors, however, call for attribution of a personal trait or external cause; the observer thereby acquires an understanding of the actor and/or situation—one that may create a lasting impression about the other individual's disposition and/or about their social circumstances (Krech et al., 1982; Shaver, 1975, 1981).

Heider (1985) identified five main strategies for determining cause: (a) "Trying"—behavior that reflects exertion and a clear intention is usually regarded as personally caused (e.g., an educator who displays considerable expenditure of energy on a worthwhile project is likely to be judged responsible for its success); (b) "Ability"—actors are held responsible if they seem to have the requisite intellectual, physical, or social ability to produce the outcomes observed (e.g., a seemingly talented professor or teacher might be regarded as the cause of high grades in a class); (c) "Unusual" or "low consensus" behavior—surprising occurrences tend to be blamed on the actors involved; (d) "Distinctiveness"—the more frequently a person acts in a recognized way, the more likely the behavior is to be viewed as a personal characteristic (e.g., if an administrator repeatedly praises all staff members for their achievements, observers may view the commendation as indicative of the administrator's flattering or manipulatory disposition: if the administrator praises staff selectively, then an

external factor—the subordinate(s) singled out for approval—is likely to be held responsible); and (e) "Task difficulty"—a highly complex task, or one that is resisted by the social environment, is more likely to be attributed to the actor (Krech et al., 1982; Shaver, 1981). In addition, a perceiver may consider not only the reasons for an observed behavior but alternative behaviors that the actor elected to forego (Jones & Davis, 1965). Behavior that seems unconventional gives the perceiver "high correspondence"—confidence of knowing what unusual personality characteristic caused it. For example, a principal who sharply criticizes district policies at a meeting of school trustees may be confidently presumed to possess strong personal attitudes about those policies. Finally, the perceiver may search for the source of cause-effect "covariation" (Kelley, 1967)—"the causal candidate which is most closely associated historically with the event being explained" (Antaki, 1982, p. 11). Three personal and environmental elements are taken into account: (a) the observer infers personal attributions from "distinctive," or uncharacteristic, actions; (b) an explanation is derived by examining the degree of "consistency" or variation in behavior across differing circumstances; and (c) "consensual" behavior by other actors indicates to the observer that environmental factors caused the behavior.

Other factors intervene in the attribution process. Socially undesirable acts prompt strong personal attributions, and unfavorable information about behavior amplifies observers' interest in identifying causes (Wrightsman, 1977). "Personal motives" make certain kinds of observers especially subject to attributional bias: those who believe "people get what they deserve" consistently blame others for their misfortunes; perceivers who anticipate similar fates usually blame circumstances (Shaver, 1981, p. 138). Trained professionals arrive at distinctive attributions for events within their fields of expertise, probably owing to their exclusive knowledge (Frasher & Frasher, 1981). There are also interactive effects among these factors. If parents or administrators expect high-performing or high-ability teachers to succeed, subsequent high attainment is likely to reinforce opinions about the teachers' high ability whereas failure may be attributed to their lack of effort.

Our attributions often result in erroneous perceptions. Actors are conscious of environmental constraints and circumstances involved in making behavioral choices: observers, however, usually have to speculate about the actors' "circumstances, history, motives, and experiences," so they focus on and over-attribute behaviors to personal factors (Jones & Nisbett, 1971, p. 5)—"the fundamental attribution error" (Ross, 1977, p. 183).

We also engage in *self*-attribution to form impressions about ourselves.[3] Knowing our own contextual factors, we tend to "blame the environment" (e.g., unsupportive politicians) for our own behavior: observers blame our personal characteristics (e.g., political naiveté, Jones & Nisbett, 1971, p. 2). Moreover, feeling confident that external circumstances are responsible, we rarely seek confirmation of self-attributions (Olson & Ross, 1985). Further self-attribution bias arises out of our explanations of academic, vocational, and social successes and failures: We tend to blame external circumstances for our own failures but see ourselves as instrumental in successes (Shaver, 1981; Wrightsman, 1977)—an "ego-protective" distortion of attributions (Hewstone, 1983, p. 3). In this regard, Rogers (1982, p. 231) reported an observed tendency for classroom teachers to describe pupils as "odd," "disturbed," or "peculiar" when explaining their students' failures. Self-attributions also affect our behavior: attributions to "stable," or long-lasting, qualities such as ability or hard work motivate future effort; attributions to transient personal factors, such as luck, mood, or fatigue, do not (Weiner, 1979).

In summary, Krech et al. (1982, p. 715) concluded from their review of research that, where perceivers witness pleasurable or undesirable behavior, are affected by behavior, place great importance on behavior, or observe their own behavior, they "are likely to both see and believe what [they] 'want' to, and . . . attribute accordingly." Furthermore, the greater the involvement of perceivers in behavior, the more likely perception bias—and conflict of perceptions—is to arise and the stronger is its probable effect. Despite current criticism of attribution theorists' reliance on rationality, predictable bias, and human desire to investigate and process social interaction information (Hewstone, 1983; Lalljee & Abelson, 1983), attributional principles have substantial relevance for educational settings: They signal a need for open communication within faculties, staffrooms, and classrooms and help in understanding educators' and stakeholders' explanations for organizational and personal successes, failures, and behavior. And as Frasher and Frasher's (1981) novel framework indicates, attribution theory also holds important insights for educational researchers.

Implicit Personality Theories

Although observations and attributions help in the development of perceptions, the perceiver finally relies on a "set of expectations about which personality traits will be mutually associated" to form organized impressions about others (Shaver, 1981, p. 139). Two principles of perception are involved: (a) despite the incompleteness of available and observed data, the perceiver develops consistent perceptions ("forms impressions") of other persons by identifying and stressing certain perceived attributes and inventing others to complete the picture; and (b) the major determinant of impressions of persons is not external stimuli but each perceiver's personal, unstated ("implicit") theory about the traits associated with different personality types (Krech et al., 1982). The perceiver categorizes other persons primarily according to physical characteristics, temperament, intelligence, and socioeconomic standing; other attributes that are assumed to follow are conferred accordingly. The perceiver also classifies each attribute on the dimensions of goodness-badness, strength-weakness, and activity-passivity (Osgood, Suci, & Tannenbaum, 1957). Consequently, on the basis of observed behavior, the perceiver can both form a unified impression about the actor and rate that person on each of the three dimensions.

In a study of sales teams in the communications industry, Gladstein (1984) found implicit theories of leadership traits to be important in the formation of impressions among team members. Furthermore, Gladstein's "most striking finding" was that these theories had led individuals to erroneous perceptions of coworkers' personalities and, hence, to unfounded explanations of their successes or failures. Similarly, in an educational context, a junior postsecondary instructor might confer on a colleague of high status the traits of coldness and insincerity. The worker may believe that these traits are not only consistent with high status but indicative of, say, relative badness, extreme strength of personality, and high activity. Despite the instructor's scant knowledge of that colleague, an impression is formed—one that is likely to be not only partially unfounded but subject to observational error based on first impressions, attribution biases, and so on. In this way, implicit personality theory highlights the important role of educators' predetermined beliefs about personality in the formation of impressions of persons, as well as the constructed unity of those impressions.

Implications for Education

Research on perception highlights the powerful, unavoidable impact of personal perspectives on all activities—including educational practice and scholarly endeavor. The following seven generalizations highlight for educators, administrators, researchers, and readers of research reports central aspects of that impact.

1. The perception process comprises a number of stages: stimulus, sensation, comparison with an implicit theory of personality and existing perceptual knowledge, attribution of cause for personal behavior, and impression formation/amendment.
2. Perceptions, rather than objective reality, at least partially shape individual attitudes and behavior.
3. Perceptions are frequently no more than approximations of reality.
4. Individuals are incapable of recognizing and verbalizing some perceptions.
5. Perceptions tend to develop as internally consistent and temporally stable structures.
6. Perceptions are affected substantially by an extensive but recurring array of factors arising out of social and cultural experiences, organizational attributes, and personality characteristics.
7. Additional factors, such as first and recent impressions, impose on perceptions of persons.

Implications for Practitioners

Such knowledge of perception may be seen by some policy makers, administrators, or other practitioners as "a tool to predict and control human behavior" (Hochberg, 1978, p. 213). For example, by manipulating organizational structures, social relationships, and events, and by displaying appropriate personality characteristics and communicating suitable information, administrators may attempt to shape the perceptions and resultant behavior of subordinates, sponsors, and other stakeholders. Litterer (1973) and Bosetti (1973) defended this use of "administrative apparatus" to integrate discrepant perceptions and efforts in schools and other organizations. Bosetti also saw a place for administrative intervention to relieve dysfunctional stress and its symptom, rapid perceptual closure.

Alternatively, educators and others may utilize their knowledge of perception theory more personally to improve their perceptual skills and interactive behavior in order to approach their daily tasks and social interaction more fully and more accurately informed. Initially, they need to be aware of the factors that distort impressions of people and bias judgments of events. For example, perception theory informs us that the quality of educational evaluations and decisions about personnel and other matters may be enhanced if perceivers resist perceptual influences such as the halo effect, logical error, and assumed similarity. By admitting and weighing evidence that contradicts existing images of persons or events, overall perceptions can be tested, revised, and refined. The principles of implicit personality theory and attribution theory also highlight the prevalence and sources of differing and inaccurate perceptions; they remind practitioners of the importance of frank, regular, organization-wide communication and of the need to watch for bias in their own and others' causal explanations of organizational performance, individual action, and conflict. Particular attention should be paid to the reasons they use to justify their own behavior, attitudes, plans, achievements, and failures, and they need to bear in mind their potential as observers for erroneously holding others fully accountable for their behavior. The foregoing generalizations and discussion highlight other important factors for which

professors, teachers, and administrators need to be alert.

Beyond awareness, practitioners can act directly to improve the quality of their perceptions—and, therefore, the appropriateness of their behavior—in educational contexts. First, resisting undue influence from initial and recent impressions, educators can consciously and systematically consider issues, persons, and events from a variety of standpoints. In this regard, Morgan's (1986) advice for sound organizational management can inform educational personnel. Morgan argued that effective professionals possess a capacity to "read social situations," whereby they intentionally

> remain open and flexible, suspending immediate judgments whenever possible, until a more comprehensive view of the situation emerges. They are aware of the fact that new insights often arise as one reads a situation from "new angles," and that a wide and varied reading can create a wide and varied range of action possibilities. (p. 12)

Indeed, Morgan's elaboration of disparate metaphorical images for viewing organizations affords a framework of multiple, divergent perspectives that administrators and others may find helpful in their effort to view events and persons in diverse ways.

Second, educators need to test their perceptions against those of colleagues, students, administrators, and stakeholders with whom they associate; this calls for a concerted strategy of frank, nonjudgmental communication in the school, university, or boardroom setting, with regular invitations to others to express support or present contrary perspectives. As an example, Kurmey (1986, pp. 3-4) commented that school-level evaluators who "actively solicit the perceptions of those whom they are evaluating" can improve understanding, communication, and the quality of evaluations, and can relieve teachers' work stress and feelings of anxiety about supervision.

Third, practitioners need to make critical thinking a feature of their professional work lives. Even more than other professionals, educators are susceptible to the confining socialization of schooling, teacher education programs, and practicum, internship, and initial teaching experiences (Lanier & Little, 1986; Patterson & Miklos, in press; Tardif, 1985; Warren, 1985) that imprisons fledgling professionals within conventional views of education and teaching and discourages analysis and critical reflection. Strategies for fostering critical reflection in education faculties and schools have been proposed elsewhere (Johnson, 1987). It is sufficient to note here that academics, policy makers, teachers, and administrators must bear individual responsibilities for exposing, discussing with colleagues, and habitually questioning existing educational values and theories and the instructional practices to which they subscribe.

Implications for Researchers

Perceptions also impose on educational inquiry in two critical ways. First, researchers draw heavily on the perceptions of individuals associated with educational organizations to learn about life in those contexts, searching either for proxy measures of an objective reality or for expressions of disparate realities for individuals. In either respect, perceptions have much to reveal about meaning and behavior in educational situations. At the same time, the perceptions gained from selected informants afford only partial views of the reality perceived by participants—perceptions prejudiced by all the prior social and cultural experiences, organizational factors, personality characteristics, and other factors that bias and otherwise shape individuals' expressed attitudes and behaviors in diverse, immeasurable ways.

Second—less obvious, but equally important—is the impact of perceptions on researchers themselves. Researchers' own perceptions shape their theories about life in educational organizations, prompt them to identify and regard certain kinds of activities and ideas as worthy of study, bias their selections among available methodological techniques, and influence the data they collect from educators and other

stakeholders. In the first place, paradigmatic stances direct and limit researchers' foci of inquiry; writers such as Burrell and Morgan (1979), Owens (1982), and (in the current volume of *The Alberta Journal of Educational Research*) Jacknicke and Rowell (1987) have made this abundantly clear. One researcher, for example, may approach the study of a school or higher education setting from the standpoint of a traditional "rationalistic paradigm" or "empirical-analytic inquiry orientation." As Jacknicke and Rowell (1987) explained, "The researcher within this orientation sees a classroom existing as a set of variables, and . . . attempts to explain and predict behaviors in terms of these variables" (p. 65). A second researcher, adhering to an "interpretive" or "naturalistic" orientation, may investigate the same social situation expecting to find socially constructed, multiple realities. This researcher may accept what Owens (1982) labeled the "naturalistic-ecological hypothesis which claims that human behavior is so significantly influenced by the context in which it occurs that regularities in those contexts are often more powerful in shaping behavior than differences among the individuals present" (p. 5); or the researcher may emphasize individuality, believing that "What is true for one person is not for another. In that sense we live in different worlds" (Greenfield, 1982, p. 5). In either case, the second researcher will see realities as multiple and context-bound and will be concerned, as Jacknicke and Rowell (1987) observed, to "make explicit the expectations, attitudes, motives, and common meanings that teachers and students have in specific situations" (p. 67). A third researcher, subscribing to a "psychic prison" (Morgan, 1986), "critical theory" (Bates, 1986), or other "radical humanist" (Burrell & Morgan, 1979) orientation, may be led to explore the social metaphors and assumptions that individuals use to dominate others, express their interests, and otherwise cope in the social setting of concern. Jacknicke and Rowell (1987) noted that, with this guiding orientation, "concerns relating to epistemology, power relationships, root metaphors, politics of schooling, and other factors that interplay to construct that phenomenon we know as education need to be examined and exposed" (p. 68). Clearly, the three researchers' orientations will direct them toward different issues and divergent explanations of circumstances and events. More than this, through their emphases on and enhanced responsiveness to specific dimensions and views of organizational life, their limited foci are attended by a potential for misconstruing actions and expressions of opinion accordingly. Morgan's (1986, p. 197) discussion of the political metaphor stressed this potential for a chosen orientation to bias the perceptions of an observer: "Under the influence of a political mode of understanding everything becomes political. The analysis of interests, conflicts, and power easily gives rise to a Machiavellian interpretation that suggests that everyone is trying to outwit and outmaneuver everyone else."[4] Again, a radical humanist perspective might lead a researcher to perceive erroneously that an individual's action or statement is fundamentally an expression of domination; conversely, to ignore this perspective may also be to overlook an important explanation for that behavior or expression of opinion.

Also critical for the quality of research is the inescapable influence of every researcher's perceptions on the processes of data collection, analysis and interpretation, and subsequent reporting of results. Initially, research methodologies are frequently dictated not only by extant knowledge and needs for further study in particular areas, but also by the preferences, interests, and self-perceived capacities of individual researchers. For those who pursue quantitative studies, data collection is subject to researchers' prior judgments about aspects of phenomena that are worthy of attention. These judgments influence the design and wording of questionnaire

items and the construction of interview schedules and management of interviews. Thereby, researchers' perceptions limit the range of subsequent responses. Perceptual biases also impose on analyses of those data, including selections among available statistical and other techniques, and on the arrangement and reporting of findings. For policy analysts, perceptions direct the selection of respondents, identification of focal issues, choices among available documentary sources as well as the relative weightings and meanings assigned to various stakeholder's expressions of opinion. Other researchers, trying to minimize the confining influences of existing theoretical knowledge and personal belief, prefer to produce detailed descriptions of events, statements, and other observations in specific contexts and to allow these raw data to direct the development of theoretical frameworks for understanding life in those contexts. Even so, few would deny that the favored research perspectives of these central research "instruments" (the researchers), their selective perceptions and explorations of events, and their heightened awareness of and identification with particular informants and viewpoints all vitally affect data collection as well as subsequent efforts to ascribe meaning and significance and to extract themes from voluminous, often conflicting, data.

What can researchers do to overcome—or at least minimize—the threat of prejudice arising from reliance on their own and others' perceptions? Initially, it should be admitted that perceptions, as powerful, pervasive guides to social behavior, are valuable objects of scholarly attention. Moreover, the first issue—informants' perceptions—is of less concern to researchers who subscribe to a strictly individual conception of reality for every respondent and informant; diverse perceptions are seen as indicative of diverse realities. Even so, these researchers must have regard for the faithfulness of the perceptions obtained. The concern remains central, however, for the majority of educational researchers, each of whom searches either for an objective reality or for a realistic account of a shared reality for inhabitants of a chosen educational setting. The trustworthiness of perceptions acquired by researchers can be actively addressed in two ways. First, researchers are better prepared to deal with perceptual bias and associated difficulties if they are conversant with and watchful for the factors (discussed above) that impose heavily on informants' perceptions and, thus, upon the quality of data acquired. Second, naturalistic researchers in education and scholars in other disciplines have devised useful strategies for enhancing and establishing the trustworthiness of perceptual data; these are worthy of consideration by all who engage in research with informants and respondents. (These strategies are noted in subsequent discussion.)

The constraint caused by factors influencing researchers' own perceptions is more subtle and, hence, more difficult to manage, yet it remains a critical constraint on the quality of all educational research. This potential for perceptions to shape—or, more disturbing, to misshape—outcomes of educational research is clear. Here again researchers bear dual responsibilities—the first involving perceptual *self*-awareness. Researchers need to acquaint themselves with the factors that influence perceptions and should commit themselves to introspection and frank appraisal and disclosure of their own research orientations, methodological biases, and inherent prejudices relating to the persons, organizations, and issues associated with each research effort. Only then may we expect researchers to be equipped to investigate and report accurately on influences that focus and restrict the outcomes they present. These proposals warrant brief elaboration.

First, researchers need to consider the powerful and deleterious influences of first impressions and personal preconceptions and assumptions. As with practitioners,

they need to avoid hasty assessments and instead should view social situations, individuals, and their own research efforts expansively. Second, researchers can enhance the quality of their work by expending effort to reflect critically on, acknowledge, and, as far as possible, accommodate opposing conceptions of reality in their studies. Ideally, the confining and potentially misleading influences of particular research orientations may best be avoided by accommodating multiple perspectives within individual research initiatives. Of course, such effort to furnish readers with a range of viewpoints leads to divergent, at times competing, interpretations, explanations, and conclusions—conclusions that, as Jacknicke and Rowell (1987) observed, may be difficult to reconcile. Moreover, this approach is frequently infeasible, given time and resource constraints. At the very least, then, scholars have an obligation to reflect on and admit the limitations that chosen research orientations impose on their investigations and resultant presentations of findings. Readers of their research reports can thereby be afforded some knowledge of the nature and extent of, and biases inherent in, those contributions to knowledge. Third, each researcher should endeavor to select and employ methodological techniques most suited to the research problems identified. The temptation to accede to personal preference must be strenuously resisted, and so the need for dispassionate review of choices is vital. Fourth, each researcher should consciously and as "objectively" as possible (having regard for the perceptual influences noted earlier) subject to personal review all personal beliefs, prior conceptions, and other known biases that relate to the educational issues under investigation. Subsequently, wherever possible, these matters should be exposed to public scrutiny in final research reports. As Dimen-Schein (1977, pp. 13-14) entreated anthropologists: "We always begin [research] with preconceptions. . . . Such preconceptions organize our perceptions. . . . We must therefore make them explicit, and we must state and explain the reasons for them." Only in this way can readers judge the trustworthiness of researchers' investigative decisions and the scholarship of their contributions.

Strategies also exist to deal directly with the potentially prejudicial power of perceptions—of informants *and* researchers—on the trustworthiness of research outcomes. They have been explicated in detail in the literature and need only be mentioned here. Nevertheless, the centrality of perceptions and likelihood of perceptual nonrepresentativeness and bias in educational endeavor highlight their relevance for all research initiatives. Educational writers (e.g., Bogdan & Biklen, 1982; LeCompte & Goetz, 1982; Lincoln & Guba, 1985; Merriam, 1985; Miles & Huberman, 1985; Owens, 1982; Skrtic, 1985), primarily those of a naturalistic bent, have advocated strategies to enhance the truth value, applicability, consistency, and neutrality of research. They include triangulation, member checks, repeated/protracted observations in natural settings, and disciplined introspection (to establish "credibility"/internal validity), purposive sampling and thick descriptions (for the sake of "transferability"/external validity), and auditable disclosure and independent audit both of the research procedures ("dependability"/reliability) and of interpretations by researchers ("conformability"/objectivity). These strategies may serve different ends for different researchers. For example, as Silverman (1985) observed, Denzin's (1970, p. 472) defence of triangulation as a device for corroborating or "adjudicating between accounts" to minimize bias and promote validity involves a use of this strategy for a positivistic intent. At the same time, by adopting an interpretive orientation, collection of multiple perspectives can provide a range of individual constructions of circumstances and events and draw attention to possible sources of perceptual disparity. In view of the factors that affect individual perceptions in diverse ways, current perception theory lends

support to inclusion of these strategies, where appropriate, in both kinds of research.

Educational researchers also have much to gain from other cautions considered and techniques deployed by field research colleagues in anthropology and sociology. Not only do ethnographers and participant observers draw on a wide range of cross-validating strategies, but they adapt methodological strategies and recheck their developing interpretations as research needs demand (Wolcott, 1982, 1985; Van Maanen, 1979). As Wax (1971, p. 10) remarked, researchers who find themselves constrained by "method, theory, or technique" need to "slip through the bars and try to find out what is really going on"—methodological creativity and versatility are emphasized. Anthropologists are also particularly conscious of the limiting and distorting power of their own culturally created values, beliefs, preconceptions, and understandings of reality, and organizing categories. To acquire emic understandings of inhabitants' experience, they strive to "place in parentheses" those personal perspectives and "to always doubt the relevance of spontaneous moments of recognition" that may spring from their own prior conceptions or misconceptions (McDonnell, 1979, p. 28). Indeed, prior conceptions are widely recognized as a blinding influence in educational and other familiar settings (Wolcott, 1985)—one that Spindler and Spindler (1982) found could be remedied only with extended time in the setting. As an additional strategy, field researchers' notes record both direct accounts of informants' perceptions and behavior and—separately—their own reflections and interpretations (Pelto, 1970; Spradley, 1979), helping to minimize the impact of the researchers' perceptions on their developing understandings and subsequent thick descriptions. Personal diaries further allow field researchers to document and explore the influence of their own perceptions on the direction and conduct of inquiry throughout the course of their investigations.

Conclusion

The pursuit of an adequate theory of perception is far from complete. Nevertheless, research effort is now being directed toward explanation of individuals' causal attributions in educational and other settings, and researchers are seeking further understanding of the significance and frequency of influence of various factors that intervene in the perceptual process. There remains a need for extensive study of the nature of perception, important factors, and extent of individual differences in perceptions in educational settings. Exploration of these avenues should provide greater insight than is now available on life in educational organizations. In the meantime, researchers who rely on their own and others' perceptions for advancing knowledge about aspects of education need to be conscious of the factors that affect those perceptions, and they have an obligation to remind readers that their findings portray reality only as it is viewed by themselves and their chosen informants. At the same time, there is consolation in the knowledge that perceptions, rather than objective reality, direct most attitudes and responses. Hence understanding of social behavior in educational settings may continue to be advanced, provided that researchers recognize and acknowledge their individual research perspectives, personal preconceptions, and preferences, and utilize appropriate research strategies to maintain trustworthiness in their inquiry.

Notes

1. There are two opposing schools of thought on the question of process. Adherents of the traditional, indirect perception approach contend that stimuli become meaningful only through a sequence of cognitive activities. Heil (1983), for example, distinguished sensory activity from the perceptual task of isolating specific features of stimuli and developing inferences or beliefs from these features, and he highlighted the latter as a subsequent, cognitive activity. On the other hand, Gibson's (1979) "ecological view" of "direct perception," defended by Michaels and Carello (1981, p. 157) and others, rejects the "impoverished input" theory and the frames of reference view of perceptual organization. They argue instead that perceivers are "active explorer[s] of the environment" seeking out meaningful information which informs them directly without a need for internal processing by perceivers; individuals admit meaningful stimuli selectively, and these trigger perceptions directly.
2. Limitations of these theoretical concepts should be acknowledged. To Neisser (1976), the selection of stimuli for perceptual processing may be a positive, rather than negative, filtering process; rather than using perceptual defences to discard items, individuals may actively choose desirable events from among those available. Furthermore, while much of the literature affords support for the concepts of constancy and threshold, Forgus and Melamed (1976) advanced rival explanations for the research findings on which these notions were founded, and Hochberg (1978) concluded from conflicting research reports that the concept of threshold in social perception is both arbitrarily defined and empirically tenuous.
3. The self-attributional process is widely thought to operate in the same general way as for other persons, although Farr and Anderson (1983) disagreed with this view.
4. Morgan's (1986) use of the dramatic "psychic prisons" metaphor also reminds us of the power of language tools, vocabulary, and communication skills to shape perceptions.

Acknowledgment

Suggestions and comments by W.H. Worth, E.A. Holdaway, C.H. Payne, and two anonymous reviewers helped considerably in the preparation of this article. They are acknowledged with gratitude.

References

Allport, F.H. (1955). *Theories of perception and the concept of structure*. New York: Wiley.

Allport, G.W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart & Winston.

Antaki, C. (1982). A brief introduction to attribution and attributional theories. In C. Antaki & C. Brewin (Eds.), *Attributions and psychological change* (pp. 3-21). London: Academic Press.

Asch, S.E. (1960). Effects of group pressure upon the modification and distortion of judgments. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (2nd ed., pp. 189-200). Evanston, IL: Row, Peterson.

Bartley, S.H. (1972). *Perception in everyday life*. New York: Harper & Row.

Bartley, S.H. (1980). *Introduction to perception*. New York: Harper & Row.

Bates, R.J. (1986, April). *The culture of administration, the process of schooling and the politics of change*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Bieri, J., Atkins, A., Briar, S., Leaman, R.L., Miller, H., & Tripodi, T. (1966). *Clinical and social judgment*. New York: Wiley.

Blake, R.R., & Ramsey, G.V. (Eds.). (1951). *Perception: An approach to personality*. New York: Ronald Press.

Blake, R.R., Ramsey, G.V., & Moran, L.J. (1951). Perceptual processes as basic to an understanding of complex behavior. In R.R. Blake & G.V. Ramsey (Eds.), *Perception: An approach to personality* (pp. 3-24). New York: Ronald Press.

Bogdan, R.C., & Biklen, S.K. (1982). *Introduction to qualitative research methods: A phenomenological approach to the social sciences*. New York: Wiley.

Bosetti, R.A. (1973). The process of perception and the administrator. *Canadian Administrator, 12*(8), 1-4.

Bruner, J.S. (1951). Personality dynamics and the process of perceiving. In R.R. Blake & G.V. Ramsey (Eds.), *Perception: An approach to personality* (pp. 121-147). New York: Ronald Press.

Brunswik, E. (1944). Distal focusing of perception: Size constancy in a representative sample of situations. *Psychological Monographs, 254.*

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis.* London: Heinemann.

Cameron, K.S., & Whetten, D.A. (1983). Organizational effectiveness: One model or another? In K.S. Cameron & D.A. Whetten (Eds.), *Organizational effectiveness: A comparison of multiple models* (pp. 1-24). New York: Academic Press.

Clement, D.E. (1978). Perceptual structure and selection. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of perception, vol. 9, Perceptual processing* (pp. 49-84). New York: Academic Press.

Dember, W.N., & Warm, J.S. (1979). *Psychology of perception* (2nd ed.). New York: Holt, Rinehart & Winston.

Denzin, N.K. (1970). *Sociological methods.* New York: McGraw-Hill.

Deutsch, M., & Gerard, H.B. (1960). A study of normative and informational social influences upon individual judgment. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (2nd ed., pp. 201-213). Evanston, IL: Row, Peterson.

Dimen-Schein, M. (1977). *The anthropological imagination.* New York: McGraw-Hill.

Ekman, P., Sorenson, E.R., & Friesen, W.V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164,* 86-88.

Epstein, W., & Rock, I. (1960). Perceptual set as an artifact of recency. *American Journal of Psychology, 73,* 214-228.

Farr, R.M., & Anderson, T. (1983). Beyond actor-observer differences in perspective: Extensions and applications. In M. Hewstone (Ed.), *Attribution theory: Social and functional extensions* (pp. 45-64). Oxford: Basil Blackwell.

Forgus, R.H., & Melamed, L.E. (1976). *Perception: A cognitive stage approach* (2nd ed.). New York: McGraw-Hill.

Frasher, J.M., & Frasher, R.S. (1981, Summer). Administrative attribution theory. *Journal of Educational Administration, 19,* 153-176.

French, W.L., Kast, F.E., & Rosenzweig, J.E. (1985). *Understanding human behavior in organizations.* New York: Harper & Row.

Gage, N.L., & Cronbach, L.J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62,* 411-422.

Geertz, C. (1979). The impact of the concept of culture on the concept of man. In D.W. McCurdy & J.P. Spradley (Eds.), *Issues in cultural anthropology: Selected readings* (pp. 35-48). Boston: Little, Brown.

Gibson, J.J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gladstein, D.L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*(4), 499-517.

Goffman, E. (1967). *Interaction ritual.* Garden City, NY: Anchor.

Greenfield, T.B. (1982). Against group mind: An anarchistic theory of education. *McGill Journal of Education, 17*(1), 3-11.

Hamachek, D. (1970). *Encounters with the self.* New York: Holt, Rinehart & Winston.

Harvey, J.H., & Weary, G. (1981). *Perspectives on attributional processes.* Dubuque, IA: Brown.

Harvey, J.H., Weary, G., & Stanley, M.A. (1985). Attribution theory and research—Still vital in the 1980s. In J.H. Harvey & G. Weary (Eds.), *Attribution: Basic issues and applications* (pp. 1-4). Orlando, FL: Academic Press.

Heider, F. (1980). On balance and attribution. In D. Görlitz (Ed.), *Perspectives on attribution research and theory: The Bielefeld symposium* (pp. 9-18). Cambridge, MA: Ballinger.

Heil, J. (1983). *Perception and cognition.* Berkeley, CA: University of California Press.

Hewstone, M. (1983). Attribution theory and common-sense explanations. In M. Hewstone (Ed.), *Attribution theory: Social and functional extensions* (pp. 1-26). Oxford: Basil Blackwell.

Hochberg, J.E. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Izard, C.E. (1971). *The face of emotion*. New York: Appleton.

Jacknicke, K.G., & Rowell, P.M. (1987). Alternative orientations for educational research. *Alberta Journal of Educational Research, 33*(1), 62-72.

Johnson, N.A. (1987). *Are programs of school-based teacher education addressing the "critical" issues?* Unpublished paper, Faculty of Education, University of Alberta.

Jones, E.E., & Davis, K.E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology, vol. 2* (pp. 219-266). New York: Academic Press.

Jones, E.E., & Nisbett, R.E. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, NJ: General Learning Press.

Kelley, H.H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15*, 192-238.

Kelly, G.A. (1955). *The psychology of personal constructs*. New York: Norton.

Kelly, J. (1974). *Organizational behavior: An existential-systems approach* (rev. ed.). Homewood, IL: Irwin.

Kelly, J. (1980). *Organizational behavior: Its data, first principles, and applications* (3rd ed.). Homewood, IL: Irwin.

Krech, D., Crutchfield, R.S., Livson, N., Wilson, W.A., & Parducci, A. (1982). *Elements of psychology* (4th ed.). New York: Knopf.

Kurmey, W. (1986). Teachers' perceptions important for effective evaluation. *Canadian School Executive, 6*(4), 3-8.

Lalljee, M., & Abelson, R.P. (1983). The organization of explanations. In M. Hewstone (Ed.), *Attribution theory: Social and functional extensions* (pp. 65-80). Oxford: Basil Blackwell.

Lanier, J.E., & Little, J.W. (1986). Research on teacher education. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 527-569). New York: Macmillan.

LeCompte, M.D., & Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52*(1), 31-60.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Litterer, J.A. (1973). *The analysis of organizations* (2nd ed.). New York: Wiley.

McDonnell, R. (1979). Understanding the encounter. In D. Turner & G.A. Smith (Eds.), *Challenging anthropology* (pp. 27-32). Toronto: McGraw-Hill Ryerson.

Mehrabian, A. (1972). *Nonverbal communication*. Chicago: Aldine-Atherton.

Merriam, S.B. (1985). The case study in educational research: A review of selected literature. *Journal of Educational Thought, 19*(3), 204-217.

Michaels, C.F., & Carello, C. (1981). *Direct perception*. Englewood Cliffs, NJ: Prentice-Hall.

Miles, M.B., & Huberman, A.M. (1985). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.

Moates, D.R., & Schumacher, G.M. (1980). *An introduction to cognitive psychology*. Belmont, CA: Wadsworth.

Morgan, G. (1986). *Images of organization*. Beverly Hills, CA: Sage.

Murphy, R.F. (1979). *An overture to social anthropology*. Englewood Cliffs, NJ: Prentice-Hall.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.

Olson, J.M., & Ross, M. (1985). Attribution research: Past contributions, current trends, and future prospects. In J.H. Harvey & G. Weary (Eds.), *Attribution: Basic issues and applications* (pp. 283-311). Orlando, FL: Academic Press.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Owens, R.G. (1982). Methodological rigor in naturalistic inquiry: Some issues and answers. *Educational Administration Quarterly, 18*(2), 1-21.

Patterson, R.S., & Miklos, E. (1987). It's time for change in teacher education: Real change this time! In L. Stewin & S. McCann (Eds.), *Contemporary educational issues: A Canadian mosaic* (pp. 522-535). Toronto: Copp Clark Pitman.

Pelto, R.J. (1970). *Anthropological research: The structure of inquiry*. New York: Harper & Row.

Rist, R.C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 40*(3), 411-451.

Rock, I. (1975). *An introduction to perception*. New York: Macmillan.

Rogers, C.G. (1982). The contribution of attribution theory to educational research. In C. Antaki & C. Brewin (Eds.), *Attributions and psychological change* (pp. 227-245). London: Academic Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 10* (pp. 174-220). New York: Academic Press.

Schlenker, B.R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations*. Monterey, CA: Brooks/Cole.

Shapiro, J.Z., & McPherson, R.B. (1987). State board desegragation policy: An application of the problem-finding model of policy analysis. *Educational Administration Quarterly, 23*(2), 60-77.

Shaver, K.G. (1975). *An introduction to attribution processes*. Cambridge, MA: Winthrop.

Shaver, K.G. (1981). *Principles of social psychology* (2nd ed.). Cambridge, MA: Winthrop.

Silverman, D. (1985). *Qualitative methodology and sociology*. Hants, UK: Gower.

Skrtic, T.M. (1985). Doing naturalistic research into educational organizations. In Y.S. Lincoln (Ed.), *Organizational theory and inquiry: The paradigm revolution* (pp. 185-220). Beverly Hills, CA: Sage.

Spindler, G., & Spindler, L. (1982). Roger Hanker and Schönhausen: From familiar to strange and back again. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational ethnography in action* (pp. 20-43). New York: Holt, Rinehart & Winston.

Spradley, J.P. (1979). *The ethnographic interview*. New York: Holt, Rinehart & Winston.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*, 1-23.

Tagiuri, R. (1969). Person perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology, vol. 3* (2nd ed., pp. 395-449). Reading, MA: Addison-Wesley.

Tajfel, H. (1969). Social and cultural factors in perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology, vol. 3* (2nd ed., pp. 315-394). Reading, MA: Addison-Wesley.

Tardif, C. (1985). Beliefs of pre-service teachers: Implications for teacher education. *Teacher Education, 27*(October), 85-93.

Van Maanen, J. (1979). The fact of fiction in organizational ethnography. *Administrative Science Quarterly, 24*(4), 539-550.

Vivelo, F.R. (1978). *Cultural anthropology handbook*. Toronto: McGraw-Hill.

Warren, D. (1985, December). Learning from experience: History and teacher education. *Educational Researcher*, 5-12.

Wax, R. (1971). *Doing fieldwork: Warnings and advice*. Chicago: University of Chicago Press.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3-25.

Wolcott, H.F. (1982). Mirrors, models, and monitors: Educator adaptations of the ethnographic innovation. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational ethnography in action* (pp. 70-91). New York: Holt, Rinehart & Winston.

Wolcott, H.F. (1985). On ethnographic intent. *Educational Administration Quarterly, 21*(3), 187-203.

Wrightsman, L.S. (1977). *Social psychology* (2nd ed.). Monterey, CA: Brooks/Cole.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 498-498*

Walter H. Worth

Reflections

Educational systems in industrialized nations are being confronted, often simultaneously, with four kinds of *scarcity*: scarcity of resources, scarcity of students, scarcity of authority and influence, and scarcity of public confidence. Emerging evidence suggests that this condition will be a *long-term* one which requires *different* responses from those to which conditions of growth have habituated us.

The challenge of scarcity to educators is to view it as an opportunity for renewal at every level of operation from classroom to council chamber, and in every sector from early childhood through adult education. For scarcity creates the need for change—to clarify purposes and to determine what to keep doing, what to do differently, what to stop doing, what to start doing. Momentum is generated for reforms that would otherwise be impossible to implement. The outcome can be a more coherent match between purposes, programs and resources; so that the resources support the programs, the programs achieve the purposes, and the purposes mobilize the resources.

This process of renewal highlights the importance of gaining and using better knowledge about what policies and practices help to achieve the desired, frequently competing, objectives of effectiveness, efficiency and equity. And bettering knowledge is what research is all about. In a time of growth, however, the impact of research was not immediate or direct. Will its impact be any different in a time of scarcity? Perhaps.

Perhaps ... if research addresses the pervasive scarcity-inspired issues of the day with a view to sharpening perceptions, stimulating discussion and encouraging questioning. Among these issues are: rethinking the purposes of schooling and education, assessing the meaning and measures of performance for individuals, programs and organizations; examining the rationale underlying current policies and practices; identifying emerging needs and responses to them.

Perhaps ... if research also enhances the "know-how" of practitioners and policy-makers. For both groups, the use of research is optional and their major motivation is to be successful. They will use research when it helps them understand and cope with their scarcity-related problems. Fortunately, our relatively new-found skills in evaluation research, naturalistic methods and policy studies put us in a stronger position than in the past to gain and offer insights and information about such context-specific concerns.

Perhaps the gap between research and practice in education, and between research and policy, can be closed—at least a little. Scarcity gives us a good opportunity to try. The extent of our success will probably be reflected in the support accorded research in the future.

Reprinted from The Editor's Page, *The Alberta Journal of Educational Research*, 31(2), 87, 1985.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 499-499*

T.O. Maguire
Editor, March 1989-June 1991

Educational Problems Revisited

In looking over the back issues of *AJER*, I was struck by the wide variety of topics covered in the papers. There is almost a sense that each generation of scholars has to revisit the central problems of education, and that there is very little memory in the system.

There are several reasons why George Buck's article "M.E. LaZerte: Pioneer educational innovator" is my choice. First, LaZerte was not only an educational innovator, he was a leading educational researcher. He and those who followed him were great believers in the idea that educational innovations of all sorts should be put to empirical test, that educational practice can be improved by carefully designed research, and that the onus is on the innovator to demonstrate the efficacy of an innovation. Beginning researchers could learn much about the merits of creativity, planning, and skepticism by following the path of LaZerte.

The second reason for choosing this article is the hope that it will remind people that we have a significant research heritage. Here was a person concerned with provincial educational matters who actually was a world-class educational scholar. I suspect that were his roots in mid-America, LaZerte would have had the recognition of Thorndike or Pressey.

Finally, when I first read the article, I was impressed by its careful scholarship and clear writing style. Recent rereading has given me no reason to change my mind. Indeed these are characteristics of all of George Buck's work. It was my hope that this article would be an example for other educational researchers. It was certainly in LaZerte tradition.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 500-510*

GEORGE H. BUCK
University of Alberta

M.E. LaZerte: Pioneer Educational Innovator

M.E. LaZerte was a prominent educator in Alberta during the first half of this century. This article describes LaZerte's learning theories and several instructional devices he created and tested. Their merits and weaknesses are discussed from LaZerte's viewpoint, as well as their relationship to general instructional theory and pedagogy. The article also compares LaZerte's innovations and learning theories to other contributors, including those of the prominent American psychologist B.F. Skinner. LaZerte's findings from his instructional devices may have relevance in assessing the efficacy of microcomputers used for instruction.

It has been over 35 years since Milton Ezra LaZerte (1885-1975) held an official position in the University of Alberta. Why should his name be remembered and why should anyone still consider the merits and the findings of his research? Although LaZerte began his career in education in rural Alberta, he did not conform to the common yet erroneous notion that most rural teachers, especially those in Western Canada, possessed pedagogical skills inferior to those of teachers in the large centers of central and eastern Canada and the United States. He not only excelled as a teacher, he also demonstrated his skills in both educational innovation and administration early in his career (Chalmers, 1978).

LaZerte was the first director of the School of Education at the University of Alberta during the 1930s, and he later became the first Dean of that university's Faculty of Education; this was the first faculty in any provincial Canadian university that prepared all teachers for that province (Johns, 1981, pp. 192-193). Immediately following World War I and well before obtaining responsible administrative positions, LaZerte addressed the problems encountered by many teachers attempting to teach difficult concepts to large and often ungraded classes. A believer in scientific research methods as well as innovative experimentation, LaZerte realized that many of the problems experienced by teachers as well as students might be alleviated in part by instructional materials designed to be used by an individual student without constant supervision (LaZerte, 1922, p. 30). It is significant to note that LaZerte began his work in this area several years before the so-called "progressive" movement took root in North American education (Charyk, 1974, pp. 277-278; LaZerte, 1945, p. 12). As early as 1922, LaZerte noted a need for such instructional materials:

Reprinted from *The Alberta Journal of Educational Research*, 35(2), 112-122, 1989.

> There is little apparatus in our schools to assist the teacher in giving a practical setting to the work that covers the field from group counting in primary years to the development of the angle-sum theorem in grade VIII geometry. (p. 30)

During the past 30 years, many individuals and groups from various disciplines, particularly outside education, have attempted either to "improve" or to "save" North American education by means of various teaching and testing machines. Certain American psychologists made the claim in the mid-1950s that no educators had even considered, much less experimented with, automated teaching (Skinner, 1954). Clearly they were unaware of the work and publications of LaZerte. The behaviorist B.F. Skinner developed what he called "teaching machines" that he felt would release individual teachers from boring and onerous tasks and enable them to spend their time doing more important duties. Skinner also claimed that his devices produced a significantly higher level of achievement among his students at Harvard University (Skinner, 1983, pp. 139-140). He could not understand most educators' reluctance to commit themselves to the use of either his machines or other teaching machines (Skinner, 1984). Another psychologist, Sidney L. Pressey, claimed that he had invented the first teaching machines during the early 1920s and that no interest was shown in them by educators, even though, as he argued, they provided "great savings in material and labor" (Pressey, 1932, pp. 671-672).

For a brief period during the 1960s, interest in teaching machines seemed to wax among many educators. Consequently, several companies designed and produced a variety of devices (Hendershot, 1964). In practice, however, teaching machines did not achieve the goals which Skinner had imagined, nor did such machines see widespread use, especially in Canada (Rutherford, 1961). No one at that time seemed to know that M.E. LaZerte had experimented with and rejected the use of auto-instructional devices of his own design, using primary and secondary school children as his experimental subjects, at least 30 years before Skinner came on the scene to promote an interest in teaching machines (Charyk, 1974, pp. 277-278; LaZerte, 1933, p. 5). If LaZerte actually tried out teaching devices, what were his systems and the machines themselves like? Why was no attempt made to have them mass-produced and used in the schools in order to improve pedagogy and student learning? And how are they relevant to our age of microcomputers? The answers to these questions can be found by studying LaZerte's experiments and the machines themselves, by considering his findings, and by taking into account the mood and educational environment of the time.

Background

LaZerte received a B.Educ. degree (a graduate degree of the time, similar to the contemporary Master of Education degree) from the University of Alberta in 1926. His thesis, *A Study of the Methods Used by Elementary School Pupils in Solving Problems in Arithmetic*, contains some interesting conclusions related to learning theory. LaZerte (1926) stated that, "the ability of pupils to find correct answers to problems in arithmetic is a poor measure of their actual problem-solving ability" (p. 102). He also found that

"much problem-solving is but a slavish adherence to type methods of procedure which become associated with certain language and number situations that exist in the presented problem" (LaZerte, 1926, p. 102). He took the findings as indications that the pedagogical methods employed by many teachers were unsatisfactory because the action of ascertaining a correct answer to a mathematical problem does not indicate to the teacher whether the underlying problem-solving strategy or concept has been either learned or understood by the student. These findings were, of course, not new. The Roman orator and educator, Marcus Fabius Quintilianus, known as Quintilian, had made the same observations as early as the first century A.D. (Quintilian, 1. 1. 24-25). It is also possible that similar observations were made by others prior to Quintilian. The common findings of Quintilian and LaZerte indicate that many educators are conservative; they believe that "traditional" methods, no matter how demonstrably inappropriate and inferior, are not to be displaced by newer ones. Both LaZerte and Quintilian also concluded that repetitive instruction such as drill and practice and plain memorization do not provide the student with an adequate facility for problem solving. This conclusion is still valid, especially in the criticism of modern "task analyses" (LaZerte, 1933, pp. 3-4; Quintilian, 10. 2. 2-5). It is of interest to note that Quintilian's and LaZerte's beliefs are diametrically opposed to those of Pressey and Skinner. Both of them advocated a behavioristic approach to learning, one in which the notion of the individual student formulating concepts about the subject matter being taught is dismissed as nonsense (Skinner, 1954). Although LaZerte acknowledged that his findings were based on a limited sample, he intended to investigate the matter further (LaZerte, 1926).

LaZerte worked on his Ph.D. at the University of Chicago under Charles H. Judd, a contemporary of Edward L. Thorndike. LaZerte's dissertation *A Diagnosis of Difficulties Encountered in Solving Problems in Arithmetic*, which built on his work for his B.Educ. thesis, was completed in 1927. Judd, unlike the behaviorists, shared LaZerte's belief that it was necessary for a student to develop a concept of what was being taught, rather than simply being able to parrot it. Judd stated that, "while formal drill gives a superficial mastery of arithmetic processes, it fails utterly to give pupils that higher training in reasoning which might be derived from an effort to understand the meaning of number operations" (LaZerte, 1933, p. x). LaZerte endeavored to devise methods and apparatus, not only to evaluate students' understanding of problem-solving processes, but also to develop devices that could instruct logical problem-solving procedures so that a student might develop an understanding of the underlying concepts.

Early Innovations

Envelope Test

One of LaZerte's first methods for assessing pupils' knowledge of problem-solving procedures was the Envelope Test (LaZerte, 1933, p. 12). The subject was presented with a large envelope on which was typed a question. Two smaller envelopes, located inside, each had a different problem-solving

procedure typed on the outside. The subject was to select only one of the envelopes, the one that was believed to display the correct procedure. This in turn contained two more envelopes with further steps. After further selection, the student would be left with one envelope, inside of which was a card. The card contained the last step of the problem on one side and a message on the other. Only one of the cards in the envelope array stated "This is the correct card"; the others stated, "You are wrong. Go back to the beginning and try again" (LaZerte, 1933, pp. 12-13). Figure 1 illustrates the hierarchical nature of the envelope test.

LaZerte noted that the number of both cards and envelopes was altered to accommodate questions of different complexity. In addition, he stated that some sequences possessed intermediate blank cards or envelopes which indicated that the examinee was on the wrong track. This modification was intended to save time, serving to inform the student of an erroneous choice early in the procedure. Although a blank card or envelope informed the examinee that an error had been made, it did not provide any corrective hint or cue (LaZerte, 1933, p. 13). In this manner, the Envelope Test did not instruct the examinee, and the procedure tested only what the examinee knew at the time of the test. The Envelope Test possessed the capability, nevertheless, of providing instruction in problem-solving strategy, if it was desired. LaZerte had created a rudimentary branching program anticipating those found in many "scrambled books" or programmed texts produced 25 years later (Lumsdaine & Glaser, 1960).

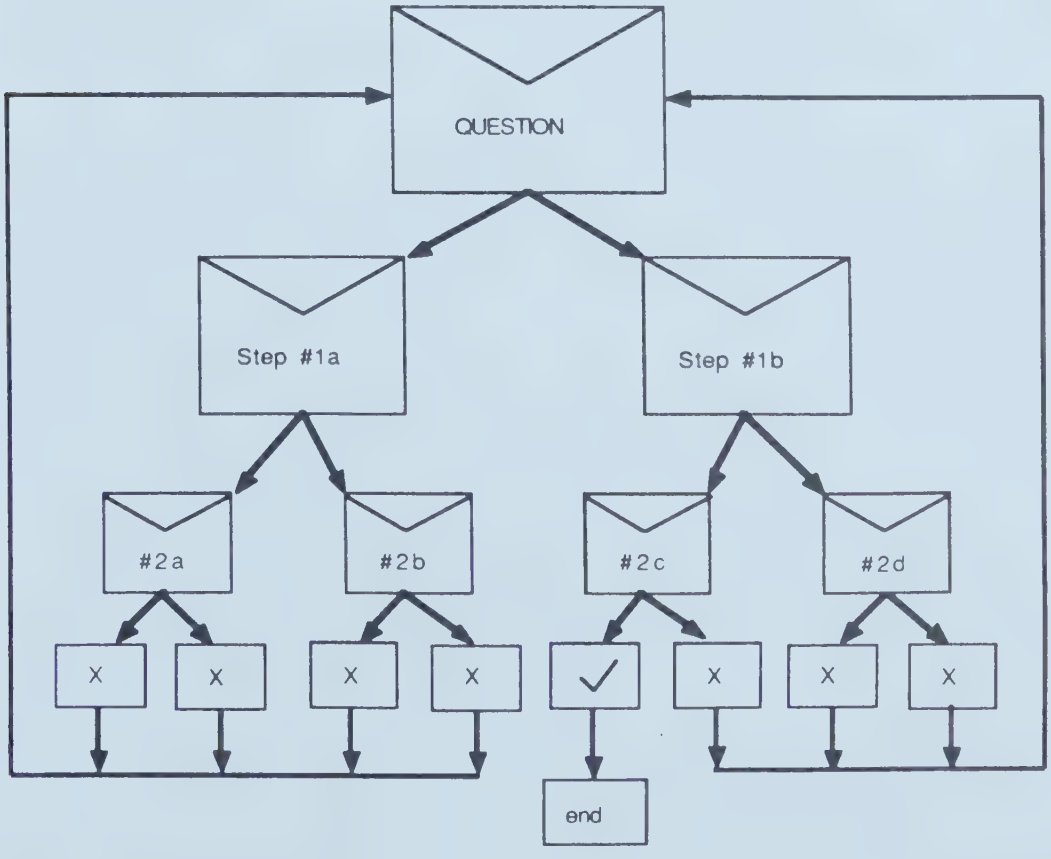


Figure 1: The hierarchical envelope test devised by LaZerte

The Problem Cylinder

There were distinct disadvantages to the Envelope Test. In order to prevent cheating, a supervisor had to be present while the test was in progress. The supervisor, as well, had to restore the cards and the envelopes after each test. Each Envelope Test, therefore, required the undivided attention of a supervisor who, if a different system were used, could have been supervising several examinations simultaneously. In addition, the relatively small number of decisions available to the examinee meant that one could possibly guess the correct sequence of envelopes as well as the correct card (LaZerte, 1933, p. 58). To alleviate these problems, LaZerte devised, and had constructed in 1930, a machine which he called the "problem cylinder" (Figure 2). Although others such as the psychologist Pressey were also developing testing machines at this time, there is no evidence in the literature or elsewhere to indicate that LaZerte's work was anything but an independent endeavor (personal communications with H. T. Coutts, 1987). There is no similarity in the concepts behind or the design of the machines or the learning theory implicit in its structure.

LaZerte's apparatus consisted of a wood and brass frame which housed five laminated wooden cylinders. The cylinders were constructed so that a keyway was present throughout their length. In addition, the center laminations were designed so that the cylinder could be rotated around a key if it was positioned in the center of the cylinder. Each cylinder possesses five brass card holders, evenly spaced about the circumference. The card holders are intended to hold segments of equations or portions of a problem-solving strategy. In between each card holder is a small hole into which one can insert a brass pin located in a bar running along the top of the frame. The pin is intended to hold the particular cylinder in the selected position. The cylinders are mounted horizontally onto a tubular brass shaft which also has five steel keys fastened to it in a linear fashion. The keys are arranged so that they are normally located within the center of each cylinder, thus permitting each cylinder to rotate freely and independently. A square brass rod, which passes through a squared hole in the frame, is attached to one end of the shaft. The squared rod prevents the shaft from rotating so that the keys on the shaft will always be positioned along the top of the shaft. If all the cylinders are rotated on the shaft so that their keyways are aligned with the keys, the shaft can be slid one and a half inches to the right. This action signifies to the user that the problem presented has been solved correctly (LaZerte, 1933, p. 59, and author's observations). The operation of the problem cylinder is similar to cylindrical bicycle locks that are used at the present time. When the correct numbers are selected on the cylinders of the lock, the keyways of the cylinders are aligned, allowing the release of a keyed shaft within the lock, thus enabling the chain or wire locking the bicycle to be removed.

A total of 3,125 arrangements of the cylinders is possible, so that the probability of an individual guessing the correct position of all five cylinders is extremely unlikely. In operation, the subject was presented with a mathematical problem and was then instructed to solve it, starting with the left-

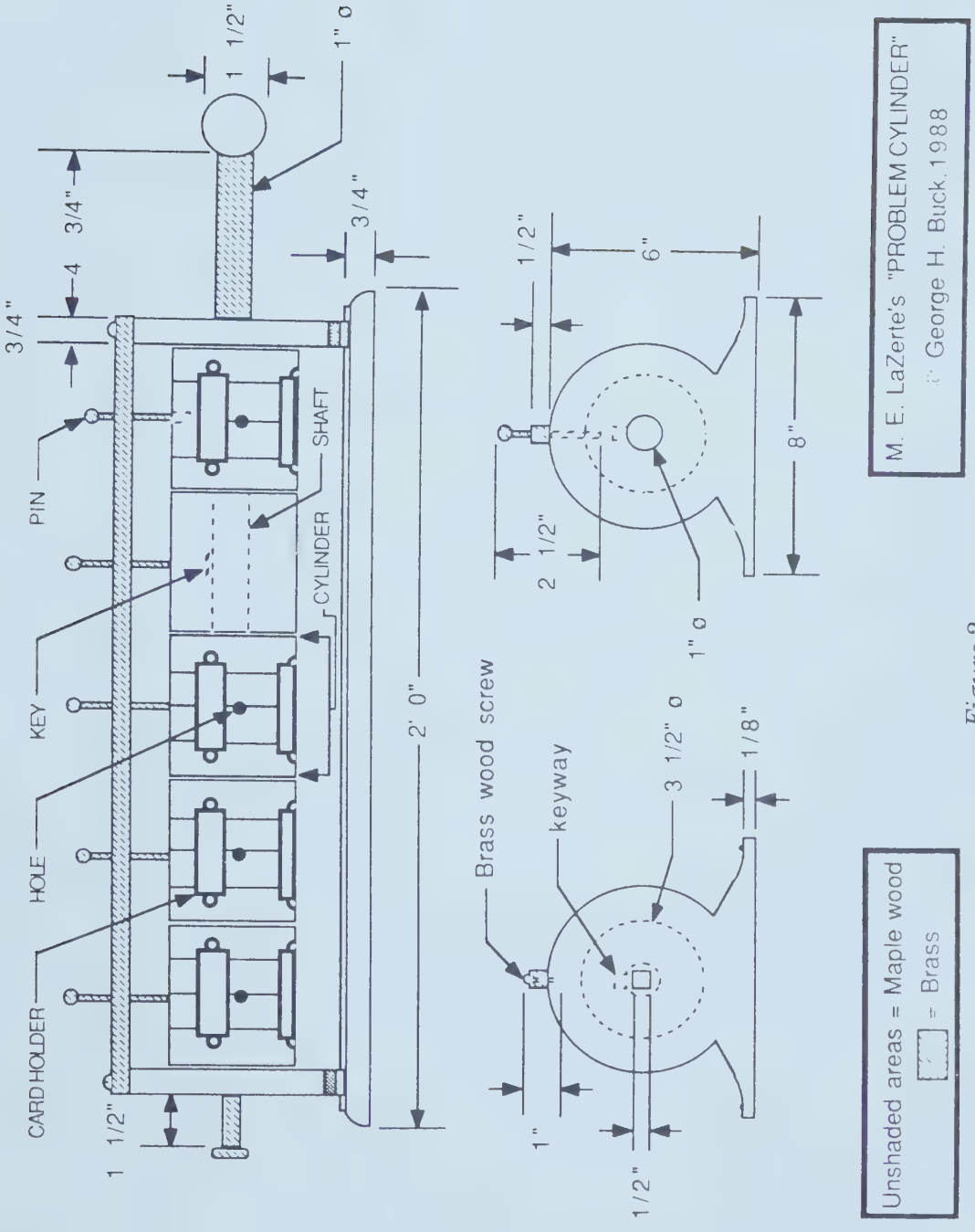

CARD HOLDER
HOLE
KEY
PIN
SHAFT
CYLINDER
1 1/2"
4
3/4"
3/4"
1" o
1 1/2"
2' 0"
Brass wood screw
keyway
1"
1/2"
3 1/2" o
1/8"
1" o
2 1/2"
1/2"
6"
8"
M. E. LaZerte's "PROBLEM CYLINDER"
© George H. Buck, 1988
Unshaded areas = Maple wood
= Brass

Figure 2

most cylinder and proceeding to the right. Every cylinder held five cards, each of which displayed a part of a problem solving strategy. The subject was to select what was believed to be the correct strategy among the five cards on each cylinder. This was accomplished by the subject rotating each cylinder so that the desired information faced him or her. After each selection was made, the subject was supposed to insert a brass pin into a hole in the top of the cylinder to prevent it from being moved while the next selection was undertaken. When it was felt that all five cylinders were in the correct position, the subject was to pull on the shaft. If the shaft did not slide to the right, one or more cylinders were improperly positioned, and this was an indication to the subject that a mistake in the steps of solving the problem had been made and that he or she should try again.

LaZerte (1933) found that the problem cylinder was intriguing to students and was highly motivational. He stated that, "all subjects were very much interested in the problem presented to them. After the experiment was completed many requested to continue their efforts. Motivation was strong throughout" (p. 60). In a fashion similar to that of the Envelope Test, the problem cylinder could be arranged so that fewer than five cylinders could be used. In such cases, those cylinders not in use would not have cards inserted in their card holders. In addition, the cylinders not in use would be pinned in the unlocked position.

Unlike the Envelope Test, however, the problem cylinder could be used without constant supervision because there was no easy way in which the subject could cheat or sabotage the machine. Besides being a testing device, the problem cylinder could be used to demonstrate logic or the logical analysis of a problem. The question noted at the outset of this article comes to mind: Why was no attempt made to have the problem cylinder mass-produced and used in the schools in order to improve pedagogy and student learning? The answer is provided by LaZerte's findings. Although high motivation and interest were attributed to the problem cylinder in a kind of "halo" effect, LaZerte (1933) also discovered that the skill required to understand and to manipulate the apparatus added to the difficulty of the problem being solved: "It is apparent that the task of solving the problems on the cylinder is much more difficult than when presented in the usual form" (p. 65). Data gathered from comparison groups verified his findings (LaZerte, 1933, pp. 63-67). It seems probable, therefore, that if LaZerte's problem cylinder were to be used successfully as a "teaching machine," instructions and practice exercises would have to be provided so that the manipulation of the machine would not add to the difficulty of the problem presented. Thus much time and effort would have to be expended on training in the use of the machine. Mutatis mutandis, the same criticism may be applied to microcomputers used for instruction. In addition to being familiar with the physical hardware such as the keyboard and peripheral devices such as a mouse, the computer user must also know the peculiarities of the programs used. It is unlikely, therefore, that the user will be able to give full and undivided attention to the primary task until some time is spent in learning about the machine and the program being used. This phenomenon

has been experienced by most microcomputer users, including some who describe the frustration caused by the problem (Vargas, 1986). Given this drawback, LaZerte concluded that the use of a teacher was more efficient and more effective than the use of his problem cylinder.

While the problem cylinder might appear to be primitive by current technological standards, it must be kept in mind that in 1930 the areas of instructional devices and electronics were both extremely limited in their development. In addition, most small school boards then in existence possessed limited funds and no electricity; it therefore seems unlikely that technologically complex and expensive apparatus would or could have been purchased by these school boards. Had his experimental studies shown the problem cylinder to be a successful teaching device, or if LaZerte, like Skinner 25 years later, had decided to market the device despite indications that it would not be successful in the education market, the problem cylinder could probably have been manufactured inexpensively because of the small number of parts. If the brass parts were replaced by steel or aluminum, and the wooden parts replaced with Bakelite, the cost could have been reduced even further.

Later Innovations

Activity Oriented Mathematics

Between the mid-1930s and the mid-1940s, LaZerte's time was primarily consumed by administrative obligations, but his interest in developing testing and instructional devices did not disappear. By the end of World War II, LaZerte had been testing new "activity" mathematics programs in the University of Alberta's demonstration school (LaZerte, 1945, p. 12). The programs, intended for primary grades, entailed considerable laboratory time for each pupil. LaZerte (1945) noted that, "the pupils spent 60% of the time devoted to arithmetic in the laboratory" (p. 12). In addition, the program was intended to be primarily individualized instruction where the students learned largely on their own and the teacher supervised. In this case, "individualized instruction" did not refer to the following of a rote program individually. It was expected that each student would perform the experiments outlined and discover for himself or herself the concept or concepts involved (LaZerte, 1951, p. 1). The laboratory work consisted of the use of instructional aids that were intended to facilitate the concrete representation of concepts, as well as to facilitate problem solving. The apparatus included bead frames, various number boards and charts, special colored disks, cut-outs, and rudimentary measuring devices (LaZerte, 1945, p. 12). While it produced favorable results, LaZerte noted that this "activity program" was similar to the methods he had used prior to 1930, "before the progressive movement reached Alberta" (LaZerte, 1945, p. 12). This work culminated in the production of the combination text and activity manual entitled *Numbers Tell Their Story*, first published in 1953.

Although the problem cylinder had been considered unsatisfactory for individualized instruction, LaZerte continued to believe that instructional apparatus could be useful in certain situations. LaZerte (1953) stated that,

"the emphasis on the use of apparatus and equipment should prove of inestimable value to the teacher in the rural and ungraded school" (p. ii). In many such schools, the teacher was responsible for several grades simultaneously, so the time that the teacher could spend with each grade was extremely limited. Although Alberta had embarked on a program of eliminating one-room rural schools through centralization during World War II, the process took many years to complete. LaZerte's observations continued to be valid for several years (Alberta Department of Education, *Annual Reports*, 1939-1965). The use of the *Numbers Tell Their Story* program appears to be ideally suited for such environments. LaZerte (1953) noted that "if he [the student] is to profit from the course, each pupil must do all the work himself. The teacher must not teach demonstration lessons. It is assumed that all pupils will have some equipment in their hands while the class is at work" (p. ii).

It should be recalled that LaZerte's desire for such an individualized program may be observed as early as 1922 (LaZerte, 1922, p. 30). At that time, there were some prominent psychologists who believed that practical and concrete methods should be used in the instruction of concepts. Thorndike (1923) proposed three criteria for a successful program:

> (1) Provide enough actual experiences before asking the pupil to understand and use an abstract or general idea. (2) Develop such ideas gradually, not attempting to give complete and perfect ideas all at once. (3) Develop such ideas so far as possible from experiences which will be valuable to the pupil in and of themselves. (p. 179)

LaZerte's (1953) *Numbers Tell Their Story* program appears to have met all three of Thorndike's criteria. To be sure, much of the material mentioned above, such as bead frames and felt cut-outs, were already available from school suppliers. Indeed, some of the items are still available (Primary School Supplies, 1987). Some of the apparatus was designed by LaZerte, however, and was not already available. Such equipment included perception cards (domino-like flash cards); number board (a wooden tray which could hold colored blocks in order to actualize addition and subtraction); place value frame (a shallow cardboard or wooden box marked in thirds, with nine slots in each third designed to hold colored disks); and the hundred board (a 23-inch square board, marked in quadrants with 25 hooks or nails arranged in rows in each quadrant, to illustrate submultiples of 100) (LaZerte, 1953; LaZerte, Dey, & Svidal, 1959). Originally, if one desired, complete sets of equipment could be ordered from the publisher (LaZerte, 1953, p. iv). In addition, concise instructions were provided on how to fabricate all the equipment mentioned in the book. As LaZerte (1953) noted, "if a school feels that it cannot undertake the expense of one set per child, the laboratory equipment is of such a simple nature that much of it can be reproduced in the school itself, at little cost" (p. iv). LaZerte's economical laboratory program was unlikely to be rejected because of its costly equipment, a consideration that did not seem to occur to Skinner and others who promoted expensive teaching machines and equally expensive instructional courseware.

Although prepared laboratory equipment packages were not offered from the publisher in later editions of *Numbers Tell Their Story*, the book remained in print until the mid-1960s (royalty statements in the LaZerte Papers, University of Alberta Archives). From this evidence, it appears that LaZerte's innovation of laboratory mathematics was reasonably successful, although it was eventually eclipsed by devices and curricula developed by individuals who took advantage of the mass-market distribution of publishers and manufacturers. By such means, many inappropriate and inferior devices and programs were, and still are, being foisted on educators as well as the general public.

Implications

In 1959, several years after his retirement, LaZerte wrote an article in which he foresaw little positive change in education and the repetition of previous mistakes because of a lack of innovation by educators. LaZerte (1959) believed that "too many of our colleagues are timid—afraid of change—afraid to use their professional freedom" (p. 49). LaZerte's observation is a likely explanation why various specialists in other disciplines have sought to improve education; it appears to them that little is being done within the discipline to improve it (Skinner, 1984). In addition, when it appears that the accomplishments of previous educators are being ignored or forgotten, others may fear that no progress has taken place.

The development and marketing of devices to enhance learning has continued during the last 50 years, with much impetus coming from the United States. While some of these devices have been useful in Canadian schools, others have been inappropriate because of their cost or incompatibility with the curriculum or school system (Rutherford, 1961). A review of the work of M.E. LaZerte indicates that as early as 1922, devices and materials to support learning were being developed and researched in Canada. Although some of LaZerte's research indicated that certain devices did not enhance learning in the most efficient manner, some of his devices and methods did, such as his *Numbers Tell Their Story* program, which was designed with Canadian schools in mind, particularly rural schools. LaZerte's work indicates that Canadian educators are sufficiently competent and innovative to engage in such research and development; they do not necessarily have to follow a lead or trend originating in the United States. Furthermore, LaZerte's experiments and findings suggest that there is a need for educators to evaluate critically the effectiveness of such devices and materials in the context of the local classroom before they are widely accepted for use in the field of education.

Acknowledgments

The author gratefully acknowledges the assistance provided by Dr. H.T. Coutts, former Dean of the Faculty of Education, University of Alberta, and that of Dr. S.M. Hunka of the Division of Educational Research Services, University of Alberta.

References

Alberta Department of Education. (1939-1965). *Annual reports*. Edmonton: King's/Queen's Printer.

Chalmers, J.W. (1978). *Gladly would he teach: A biography of Milton Ezra LaZerte*. Edmonton: The ATA Educational Trust.
Charyk, J.C. (1974). *Pulse of the community: Volume II of the little white schoolhouse*. Saskatoon: Western Producer Book Service.
Hendershot, C.H. (1964). *Programmed learning: A bibliography of programs and presentation devices*. Bay City, MI: Author.
Johns, W.H. (1981). *A history of the University of Alberta: 1908-1969*. Edmonton: University of Alberta Press.
LaZerte, M.E. (1922). Elementary mathematics. *The ATA Magazine Easter Annual*, 30.
LaZerte, M.E. (1926). *A study of the methods used by elementary school pupils in solving problems in arithmetic*. Edmonton: Unpublished Bachelor of Education thesis.
LaZerte, M.E. (1927). *A diagnosis of diffulties encountered in solving problems in arithmetic*. Chicago: Unpublished doctoral dissertation.
LaZerte, M.E. (1933). *The development of problem solving ability in arithmetic: A summary of investigations*. Toronto: Clarke, Irwin.
LaZerte, M.E. (1945). *I love freedom*. Edmonton: Unpublished speech notes.
LaZerte, M.E. (1951). *Arithmetic in the primary grades*. Edmonton: Unpublished speech notes.
LaZerte, M.E. (1953). *Numbers tell their story*. Toronto: Clarke, Irwin.
LaZerte, M.E. (1959, May 6-10). The road ahead. *The ATA Magazine*, 49.
LaZerte, M.E., Dey, J.D., & Svidal, R. (1959). *Numbers tell their story: Grade two teacher's manual*. Toronto: Clarke, Irwin.
Lumsdaine, A.A., & Glaser, R. (Eds.). (1960). *Teaching machines and programmed learning: A source book*. Washington, DC: National Education Association.
Pressey, S.L. (1932). A third and forth contribution toward the coming "industrial revolution" in education. *School and Society, 36*, 668-672.
Primary School Supplies. (1987). *School catalogue*. Richmond Hill, ON: Author.
Quintilian. *Institutio oratoria, I* [Training in oratory, I] (H.E. Butler, Trans., 1921). London: William Heinemann.
Rutherford, G. (Ed.). (1961). *Programmed learning and its future in Canada*. Ottawa: Canadian Teachers' Federation.
Skinner, B.F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*, 86-97.
Skinner, B.F. (1983). *A matter of consequences: Part three of an autobiography*. New York: Knopf.
Skinner B.F. (1984). The shame of American education. *American Psychologist, 39*, 947-954.
Thorndike, E.L. (1923). *The psychology of arithmetic*. New York: Macmillan.
Vargas, J.S. (1986). Instructional design flaws in computer-assisted instruction. *Phi Delta Kappan, 67*, 738-744.

The Alberta Journal of Educational Research *Vol. XL, No. 4, December 1994, 511-516*

George H. Buck
University of Alberta

George Murray Dunlop

This special issue concludes with research abstracts from the current recipients of the G.M. Dunlop awards for outstanding theses at the master's and doctoral levels.

George Murray (Pat to those who knew him) Dunlop was born early in this century in Ontario. By the time he reached high school age, his family had moved to Saskatoon, where he finished school and earned a BA from the University of Saskatchewan. Seeking his fortunes further west, Dunlop taught high school history in Medicine Hat and Calgary with considerable success. In 1927, Pat Dunlop was appointed Inspector of Schools for the Foremost district in southern Alberta. Further promotion was rapid, as he was appointed to the staff of the Edmonton Normal School in 1928 to teach psychology and history.

While at the Edmonton and Camrose Normal Schools, Dunlop completed his MA in history and began to conclude that personal upgrading of qualifications was insufficient for the discipline as a whole; there had to be ongoing academic research. This idea occurred to others as well. Obtaining a leave from the staff of the Edmonton Normal School in 1941, Major Dunlop returned in 1945 to find that the normal schools were no more, and that education was considered a legitimate discipline, now a Faculty of Education at the University of Alberta (Edmonton and Calgary campuses). Dunlop soon earned a doctorate from Teachers College, Columbia University, and began to push for the formation of a scholarly outlet for educational research in Alberta. To this end he was instrumental in laying the groundwork for the establishment of the *Alberta Journal of Educational Research.* Preferring to advance the cause of educational research in the trenches, Dunlop did not desire to be editor of the *AJER.* Until shortly before his death on July 12, 1966, G.M. Dunlop remained active both in teaching and in the promotion and fostering of educational research at the University of Alberta.

The Dunlop awards, presented by the Canadian Association of Educational Psychologists of the Canadian Society for Studies in Education (CSSE), were created in memory of Dr. Dunlop.

Research Abstracts of the 1993, 1994 Dunlop Award Winners

Effects of Teaching Statistical Laws on Reasoning About Problems Abstract: Doctoral Dissertation

Pete Kosonen
Simon Fraser University

Contemporary psychology, in the tradition of Thorndike and James, has largely been characterized by the belief that instructionally useful inferential rules are not abstract or general. In contrast, recent research supports a view that instruction in abstract rule systems can improve reasoning about ill-defined problems that abound in real life. More specifically, the research shows that instruction in statistics can influence the way people reason about events involving uncertainty in everyday life and can produce knowledge that readily transfers to types of problems outside those focused on in instruction.

Building on a study by Fong, Krantz, and Nisbett (1986), who found that training people to use the heuristic of the law of large numbers substantially enhanced their reasoning about everyday statistical problems, three experiments that investigated the effects of instructing students about this statistical heuristic were performed. The experiments were carried out with a total of 315 participants in university as well as secondary and elementary schools.

The results (a) indicate that students learned a good deal about how to reason statistically as a consequence of instruction; (b) reveal no problem format-specificity of instructional effects; and (c) suggest that typical reasoning errors were reduced by instruction in the statistical heuristic. These results stand in contrast to antiformalist views that stress domain-specificity in problem solving and lend support to the formalis view that teaching people to apply formal rules of inference can help them to reason more accurately about a variety of probabilistic events. In general, the findings bolster an optimistic view about the potential of rather limited instruction to foster valid reasoning about problems for students in university as well as in secondary and elementary schools.

Pete Kosonen is currently a school principal in Coquitlam, BC. His primary research interest is in the psychology of problem solving and transfer with a view to developing instructional programs to enhance reasoning about problems, based on finding out how people learn to think.

The Effect of Problem Text on Solving Difference-Finding Word Problems
Abstract: Master's Thesis

Ning Fan
University of Calgary

This study investigated the effect of problem text on simple arithmetic word problem solving involving difference finding between two disjoint sets. Data on correct solutions and problem solving strategies for three types of such word problems, namely COMPARE, EQUALIZE, and WON'T GET, were collected from first-grade students and analyzed in a repeated-measures analysis of variance design. The correct solution scores of the EQUALIZE and WON'T GET problems were found to be significantly higher than those of the COMPARE problems. The strategy results further revealed the dependency of strategy use on the type of problem text, in that the EQUALIZE problems were most frequently solved by using an ADD-ON strategy, and the WON'T GET problems by a MATCH strategy, whereas there was no one strategy used significantly more than others for the COMPARE problems. The use of the ADD-ON and MATCH strategies reflected the construction of a coordination between two mental number lines while the problem representations were being built. These results suggested that the EQUALIZE and WON'T GET problem texts facilitated the coordination between the two number lines, but the problem text of COMPARE did not facilitate the coordination which was crucial for solving the difference finding problems. This study shows the value of incorporating a cognitive development framework with the domain-specific mathematical and linguistic models of word problem solving.

Ning Fan received his Bachelor of Education degree from Beijing Normal University in 1985, and Master of Science degree from University of Calgary in 1993. He is currently studying in a PhD program in University of Pittsburgh.

A Cognitive Developmental Analysis of the Interpretation of Family Stories by Adolescents and Preadolescents
Abstract: Master's Thesis

Diane J. Salter
University of Calgary

This exploratory study investigated cognitive developmental changes in the understanding of family stories by children and adolescents. A neo-Piagetian theoretical framework was used to assess the developmental progression in the meaning that subjects attributed to their family story.

Children and adolescents (aged 10, 12, 14, and 18 years) were asked to relate a "family story." The meaning that the story held for them was explored by interviewing each child about his or her understanding and interpretation of the story content. The results of the qualitative and statistical analysis demonstrated that children's understanding of the meaning of the story changed according to a developmental schedule. Responses were organized into four categories or levels of response. These levels were consistent with theory in cognitive development proposing developmental increases in structural complexity. A qualitative difference was noted between the responses given at age 10 and the responses given at age 12. No significant difference in response was noted between age 12 and age 14. A statistically significant difference occurred from age 14 to age 18.

A linear developmental progression was noted, with subjects at age 10 scoring significantly lower than all other groups and subjects at age 18 significantly higher. The younger child focused mainly on the action and interpreted the story as primarily a patterning of events motivated by feelings, desires, and beliefs, whereas the older child recognized that the story pattern contained a message that could be generalized to others. The older child, thereby, voiced his or her understanding of the world by interpreting the meaning of the story.

Diane Salter is currently completing her doctorate at the Ontario Institute for Studies in Education in the Department of Cognitive Psychology. Her primary research interest is adolescent understanding of intention in oral and written history.

Effects of Conflict and Knowledge-building Approach on Conceptual Change Abstract: Doctoral Dissertation

Carol K.K. Chan
Ontario Institute for Studies in Education

This study examined knowledge construction and conceptual change in the context of how high-school students process contradictory information in the domain of biological evolution. Investigations using a conflict-based approach to foster conceptual change have yielded equivocal findings. Based on the conceptual framework of constructive activity in learning, this study aimed to investigate whether the contrasting approaches of direct assimilation and knowledge building differentially affected conceptual change, and whether the effects of conflict were mediated by knowledge-processing activity. A related objective was to examine whether peer interaction fostered conceptual change when students were presented with new concepts which contradicted their joint understanding.

In order to assess individual differences in processing contradictory information, a computer-based connectionist methodology was developed that would allow the experimenter to select new information contingent upon students' expressed beliefs. The sample consisted of 108 students in grades 9 and 12 randomly assigned to four conditions: individual assimilation, individual conflict, peer assimilation, and peer conflict. Depending on the conditions, students were asked to think aloud or discuss with their peers eight scientifically valid statements presented in the order which either maximized or minimized conflict. Verbalizations were tape-recorded for subsequent analyses of knowledge-processing activity. Several measures of posttest conceptual change were also obtained.

Protocols were coded for five different levels of knowledge-processing activity. Two major approaches were identified: direct assimilation which involves fitting new information with what is already known, versus knowledge building which involves treating new information as something that needs to be explained. A path analysis showed that only knowledge-processing activity exerted a direct effect and mediated the effects of conflict on conceptual change. Merely confronting students with contradictory information was not sufficient unless students actively carried out a knowledge-building approach to deal with cognitive conflict. No main effects for peer interactions were obtained; the interaction effects, however, indicated that older students in the conflict group benefitted more from peer learning. Knowledge-building as a mediator of conflict in conceptual change helps explicate previous equivocal findings and highlights the importance of students' constructive activity in advancing their knowledge.

Carol Chan completed her PhD at the Ontario Institute for Studies in Education under the supervision of Dr. Carl Bereiter. Her research interests include text processing, problem solving, and collaborative learning.

Children's Perceptions of the Learning Process
Abstract: Master's Thesis

Gillian Bickerton
University of British Columbia

Although there is extensive literature on the learning process, what is apparent is the relative absence of any in-depth inquiry into what learning itself means to children and where they believe learning "comes from." The objective of the study was to determine if there is a developmental sequence in children's understanding of learning, such that children of the same age think in a similar fashion about learning and there is increasing complexity in understanding with age. A neo-Piagetian model of intellectual development (Case, 1985, 1992) was used as a theoretical framework. Children aged 6 to 12 years were interviewed about their understanding of the meaning of learning.

The study indicated significant age-related differences in level of understanding of the meaning of learning; they were hierarchical in nature supporting the theoretical predictions of a developmental sequence postulated by Case (1985, 1992). children progressed from a simple conception of learning defined in terms of intentionality, in other words, learning as a behavioral act combined with an internal feeling or judgment state, to more complex notions of intentional behavior as they relate to the process of learning. An interpretive understanding of the learning process characterized the 12-year-olds' responses; that is, responses were more psychological in nature with recognition of "states of mind" as important in learning.

The complexity of the "source" responses also increased with age. In the early stages there was a clear distinction between external sources (learning located in an action or "learning agent") and internal sources (learning located within the physical or psychological self). For older children there was an awareness of learning taking place in a sequential manner from external to internal. Interrelatedness of the two sources was generally recognized by 12 years of age.

By revealing common age-typical patterns of understanding of the learning process, this study suggests that educational methods and materials should be consistent with children's levels of conceptual development (compare Case & McKeough, 1990; Case, Sandieson, & Dennis, 1986).

If the goal in education today is to assist children in becoming independent learners, educators need to first of all understand children's conceptions of learning. This will result in more effective instructional design.

Gillian Bickerton is an experienced elementary school teacher having taught in London, Ceylon, Saskatchewan, and Langley, BC. She has been a faculty associate at Simon Fraser University and recently a learning assistance helping teacher in the Langley school district. She is currently working as a resource teacher in the same district while pursuing a doctorate at the University of British Columbia.

Index, Volume XL, 1994

Anderson, J.O. *Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency,* No. 2, 247-262.

Bachor, D.G. *Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency,* No. 2, 247-262.

Baker, H.S. *A Study of the Written Composition of a Representative Sample of Alberta Grade Four and Grade Seven Pupils,* No. 4, 397-410 (reprint).

Bateson, D. *Psychometric and Philosophic Problems in "Authentic" Assessment: Performance Tasks and Portfolios,* No. 2, 233-245.

Bateson, D.J. *Verification of a Model of Test-taking Behavior of High School Seniors,* No. 2, 195-211.

Begley, P.T. *Contemporary Images of School Leadership in Ontario, the Northwest Territories, and Western Australia,* No. 3, 319-336.

Berger, A. *The Education of Canadian Indians: An In-Depth Study of Nine Families,* No. 4, 448-456 (reprint).

Bickerton, G. *Children's Perceptions of the Learning Process.* Abstract: Master's Thesis, No. 4, 516.

Bisanz, G.L. *Studying the Development of Academic Skills in the 1990s: Implications for Assessment,* No. 2, 127-146.

Bisanz, J. *Studying the Development of Academic Skills in the 1990s: Implications for Assessment,* No. 2, 127-146.

Blase, J. *The Micropolitics of Teacher Work Involvement: Effective Principals' Impacts on Teachers,* No. 1, 67-94.

Buck, G.H. *Herbert T. Coutts and the Origins, Early Development, and Possible Future Directions of the Alberta Journal of Educational Research,* No. 1, 3-6.

Buck, G.H. *George Murray Dunlop,* No. 4, 511-516.

Buck, G.H. *Introduction: Special Issue on the 40th Anniversary of the Alberta Journal of Educational Research,* No. 4, 393-394.

Buck, G.H. *M.E. LaZerte: Pioneer Educational Innovator,* No. 4, 500-510 (reprint).

Cameron, J. *Editorial: AJER: Forty Years,* No. 1, 1-2.

Cameron, J. *Higher Education Under Attack in Alberta: Short-term Versus Long-term Consequences,* No. 3, 263-264.

Cameron, J. *Introduction: Special Issue on the 40th Anniversary of the Alberta Journal of Educational Research,* No. 4, 393-394.

Carney, R.J. *History of Education and the Rite of Passage to Teaching: The Alberta Experience 1893-1945,* No. 1, 35-56.

Chan, C.K.K. *Effects of Conflict and Knowledge-building Approach on Conceptual Change.* Abstract: Doctoral Dissertation, No. 4, 515.

Cole, A. *Problems and Paradoxes in Beginning Teacher Support: Issues Concerning School Administrators,* No. 3, 297-318.

Coutts, H.T. *A Study of the Written Composition of a Representative Sample of Alberta Grade Four and Grade Seven Pupils,* No. 4, 397-410 (reprint).

Fan, N. *The Effect of Problem Text on Solving Difference-Finding Word Problems.* Abstract: Master's Thesis, No. 4, 513.

Hunka, S. *Alberta Journal of Educational Research: 1955-1965,* No. 4, 395-396.

Gupta, L. *Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics,* No. 2, 163-193.

Haig, B., *Construct Validity and Achievement Assessment,* No. 2, 109-126.

Hattie, J. *Construct Validity and Achievement Assessment,* No. 2, 109-126.

Hersom, N. *Twenty-five Years of Research in Education: The Alberta Journal of Educational Researach 1955 to 1979,* No. 4, 458-471 (reprint).

Hodysh, H.W. *History of Education and the Rite of Passage to Teaching: The Alberta Experience 1893-1945,* No. 1, 35-56.

Hodysh, H.W. *Hersom and the Development of Educational Research,* No. 4, 457.

Hodysh, H.W. *Educational Research in Transition: An Historical Perspective on Editorial Change,* No. 4, 472-473.

Housego, B.E.J. *How Prepared Were You To Teach? Beginning Teachers Assess Their Preparedness,* No. 3, 355-373.

Idle, T. *Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations,* No. 1, 57-65.

Johnson, G.M. *Undergraduate Student Attrition: A Comparison of Students Who Withdraw and Students Who Persist*, No. 3, 337-353.
Johnson, N.A. *The Pervasive, Persuasive Power of Perceptions*, No. 4, 475-497 (reprint).
Kaspar, V. *Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations*, No. 1, 57-65.
Kirby, J.R. *Introduction: Special Issue on Cognition and Assessment*, No. 2, 105-108.
Kirby, J.R. *Measuring and Predicting Depth of Processing in Learning*, No. 2, 127-146.
Kosonen. P. *Effects of Teaching Statistical Laws on Reasoning About Problems.* Abstract: Doctoral Dissertation, No. 4, 512.
Krull, C.D. *Level of Education, Sexual Promiscuity, and AIDS*, No. 1, 7-20.
Litke, C.D. *Implementing the Modified Four-day School Week*, No. 3, 271-281.
Maguire, T. *Construct Validity and Achievement Assessment*, No. 2, 109-126.
Maguire, T.O. *Educational Problems Revisited*, No. 4, 499.
Maguire, T.O. *Evaluation of the IPI Project*, No. 4, 426-444 (reprint).
McFetridge, P.A. *Mathematics Learning in Young Children*, No. 4, 411.
McLeish, J. *Native Education in the 1970s*, No. 4, 446-447.
Miklos, E. *Individualizing Learning*, No. 4, 425.
Miklos, E. *Reflections*, No. 4, 445.
Muir, W. *Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency*, No. 2, 247-262.
Nagy, P. *Cognitive Assessment of Solutions to an Ill-structured Problem Using Forced Classification Analysis*, No. 2, 163-175.
Nelson, L.D. *Conservation of Length: Methodological Considerations*, No. 4, 412-424 (reprint).
Nesbit, J.C. *Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics*, No. 2, 163-193.
Randhawa, B.S. *Theory, Research, and Assessment of Mathematical Problem Solving*, No. 2, 213-231.
Roberts, J. *The Micropolitics of Teacher Work Involvement: Effective Principals' Impacts on Teachers*, No. 1, 67-94.
Roessingh, H. *ESL Drop Out: The Myth of Educational Equity*, No. 3, 283-296.
Rogers, W.T. *Verification of a Model of Test-taking Behavior of High School Seniors*, No. 2, 195-211.
Salter, D.J. *A Cognitive Developmental Analysis of the Interpretation of Family Stories by Adolescents and Preadolescents.* Abstract: Master's Thesis, No. 4, 514.
Sawada, D. *Conservation of Length: Methodological Considerations*, No. 4, 412-424 (reprint).
Stambaugh, S. *Getting Educated*, No. 3, 265-269.
Walker, K.D. *Notions of "Ethical" Among Senior Educational Leaders*, No. 1, 21-34.
Willoughby, T. *Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations*, No. 1, 57-65.
Wilson, R.J., *Introduction: Special Issue on Cognition and Assessment*, No. 2, 105-108.
Winne, P.H. *Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics*, No. 2, 163-193.
Wood, E. *Enhancing Adolescents' Recall of Factual Content: The Impact of Provided Versus Self-Generated Elaborations*, No. 1, 57-65.
Walsh, J. *Classroom Assessment and the Relationship to Representativeness, Accuracy, and Consistency*, No. 2, 247-262.
Watt, D. *ESL Drop Out: The Myth of Educational Equity*, No. 3, 283-296.
Woodhouse, R.A. *Measuring and Predicting Depth of Processing in Learning*, No. 2, 127-146.
Worth, W.H. *Perceptions*, No. 4, 474.
Worth, W.H. *Reflections*, No. 4, 498.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes;* any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
845 Education Centre South
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1438